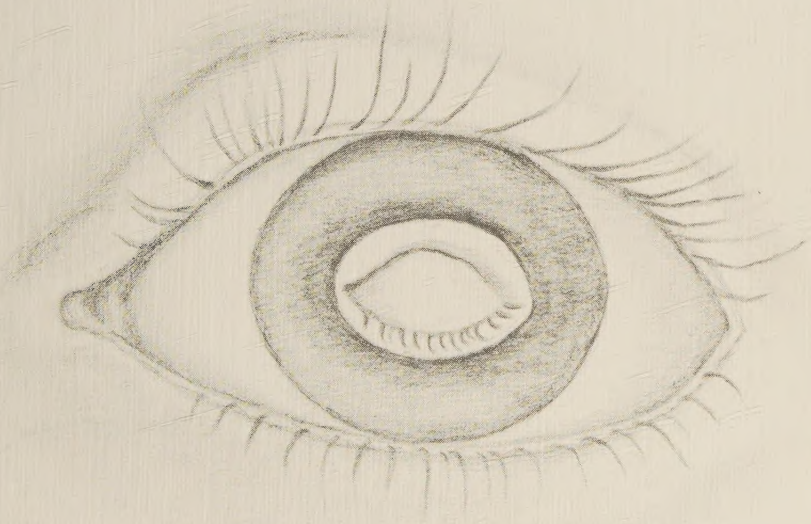
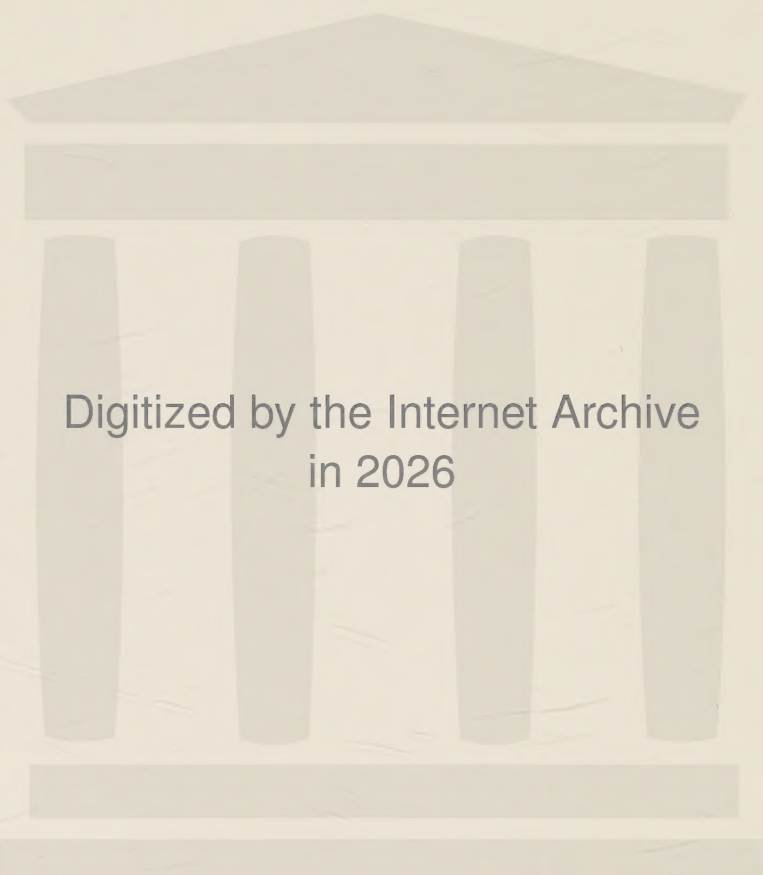


Lund 1981.

REASONS
AND
REASON-GOVERNED ACTIONS



INGMAR PERSSON



Digitized by the Internet Archive
in 2026

https://archive.org/details/IA41553210_0014

REASONS
AND
REASON-GOVERNED ACTIONS

INGMAR PERSSON



Printed in Sweden
Studentlitteratur
Lund 1981
ISBN 91-7222-360-X
ISSN 0348-5110

PREFACE

I would like firstly to thank Professor Göran Hermerén for unflagging encouragement and for useful advice in matters ranging from purely philosophical to practical ones, throughout the years I have been working on this thesis. I am also greatly indebted to Mr J.L. Mackie and Professor Don Locke, for penetrating supervision during a stimulating year in Oxford in 1979/80. My gratitude is also due to Professor R.M. Hare, Dr Anthony Kenny and Mr David Pears who, by their willingness to discuss topics related to my thesis, further enriched this year. I would like to acknowledge the generosity of the European Council for Higher Education for financing my year in Oxford. Thanks also to participants in seminars held at the University of Lund where earlier drafts of my thesis have been discussed.

Finally, I am grateful to Nicola Jackson for advice concerning my English, to Carolina Rundquist for helping me realize my ideas for the cover and to my mother for typing and proof-reading.

Ingmar Persson
Lund, December 1980

"Each of us awaits his moment in order
to propose something - anything.
He has a voice: that is enough."

E.M. Cioran
(trans. R. Howard)

TABLE OF CONTENTS

INTRODUCTION	1
1. REAL AND PHENOMENAL REASONS FOR ACTION	8
2. A CONCEPT OF ACTION	12
2.1 Outline of a Broad Concept of Action	13
2.2 Comparison of Some Theories of Action	16
2.3 Further Specification of the Broad Concept of Action	22
2.4 Some Distinctions in the Domain of Action	26
3. THE CONCEPT OF THINKING	30
3.1 Thought and Belief	30
3.2 The Content of Propositional Thinking	36
3.3 Thought and Behaviour	42
3.4 The Medium of Thought	51
4. CONDITIONALS AND POSSIBILITIES	63
4.1 An Analysis of Conditionals	63
4.2 The Spectrum of "Can"	73
5. REASONS AND WANTING	86
5.1 The Relation Between Reasons and Wanting	86
5.2 The Object of Wanting	93
5.3 The Concept of Wanting: Basic Features	101
5.4 The Concept of Wanting: Refinements	110
6. THE CONCEPT OF A REASON-GOVERNED ACTION	129
6.1 The Problem of How Reasons Govern Actions	129
6.2 Reason-governed Action: Basic Actions	140
6.3 Reason-governed Action: Complex and Non-basic Actions	151
6.4 The Domain of Application of (IBA)	157
CONCLUSIONS AND LIMITATIONS	164
NOTES	175
BIBLIOGRAPHY	186
INDEX	191

INTRODUCTION

1. The aim of this work is twofold. First, to set out necessary and sufficient conditions for someone's having a reason for action in his mind or for someone's considering or thinking of a reason for action. Second, on the basis of this account, to define the notion of a reason-governed action, i.e. the notion of somebody acting for a reason in such a way that he intentionally accomplishes both the action the reason is a reason for performing and what he calculated to achieve by doing this action. I regard it as a desideratum of such analyses that they do not appeal to action-theoretical concepts that are sui generis, irreducible; the elimination of specific action-theoretical concepts from the vocabulary of primitives, required for an adequate description of the world, obviously simplifies this vocabulary. Thus, a theory that meets the desideratum brings along at least some gain - of simplicity - even if the concepts figuring in its explanations are just as philosophically perplexing as those that are explained.

The plan of this essay is as follows.

Chapter 1: The phrase "reason for action" turns up in both explanatory and evaluative contexts. I maintain that in the former context we are interested in an agent's reasons in the sense of propositions that form contents of his thoughts ("phenomenal reasons"), whereas in the latter context we search for propositions that are true ("real reasons"). But I also contend that the reason-propositions in both contexts have a common form: that of conditionals, wherein the antecedent is constituted by a description of the action-type the reason is a reason for performing an instance of. As already hinted at, it is phenomenal reasons that will be at the centre of attention, although the topic of real reasons is not allowed to drop out entirely, since a theory of phenomenal reasons can be deemed satisfactory only if it is at least indicated how it could be extended to cover real reasons as well.

Chapter 2: I here define a broad concept of action which is logically independent of such "intentional" concepts as those of a reason for action, a want, an intention etc. Roughly, someone acts in this sense just in case some fact of change pertaining to him causes a state of affairs to come to obtain. It is important to have access to a concept of action that is independent of the concept of a reason, since one can then, without incurring any charge of circularity, appeal to this action-concept in the clarification of reasons for action.

Chapter 3: The topic of investigation here is aspects of the concept of thought that are relevant for my theory of reason-governed action, a theory

which presupposes that thoughts play a causal role in the genesis of such actions. I defend the traditional doctrine that a thought is a mental episode and so satisfies one requirement for being a cause of observable behaviour: that of being an existent distinct from the effect. But I also emphasize that the traditional doctrine goes wrong in taking propositional thinking to involve an act - of "assenting" or "endorsing" -, a thesis which, were it true, would create difficulties for an attempt to define an intentional action in terms of the causal power of propositional thinking. In this chapter I also examine (a) the intensionality of the context "thinks that ..." - which is necessary for circumscribing exactly what is intentionally done - and (b) the medium of thinking, something which is essential for an assessment of the phenomenological plausibility of my account of reason-governed actions.

Chapter 4: Here I explicate the conditional constructions "if (or only if) -then" which figure in my standard-formula of reasons. The account proposed is of the "connection"-type: putting forward a conditional is suggested to be equivalent to asserting that the consequent is deducible from the antecedent, plus, in most cases, certain implicit background-assumptions. Dealing with conditionals involves distinguishing some different uses of "can" - one of which is presupposed by the antecedent of the conditional. The job of reviewing different uses of "can" is carried further in the second section of this chapter, because clarification of these uses will be found necessary for understanding the central notions of wanting and acting intentionally.

Chapter 5: The set of conditions necessary and sufficient for somebody's having a phenomenal reason for action is now completed by the clause that the agent must want the state of affairs described in the consequent of the reason-formula. A dispositional account of wanting is then presented and defended: to want to bring about p is analyzed, roughly, as to be in a state such that one's thinking that one might have the ability to cause p tends to cause - perhaps because p is linked to other states of affairs in a reason-conditional - a state of affairs to obtain, because one thinks that it is the state of affairs one has ascribed to oneself the ability to cause, viz. p. According to this, a want to cause p is not a disposition to cause what is in fact p, but what the agent thinks is p. It follows that the agent thinks that he is capable of causing p: this is implied by his thinking that he is causing p. It is this ability-thought - perhaps supported by other states of affairs it is linked to in reason-conditionals - which causes what is thought to be an exercise of the ability.

Chapter 6: When somebody acts for a reason (in the sense characterized above) he brings about the materialization of one state of affairs in an

intentional basic action and the materialization of at least one other state of affairs in an intentional non-basic action, by bringing about the reality of the first state of affairs. That somebody causes p in an intentional basic action is defined as that he is in a state such that his thinking that he can cause p causes - perhaps due to its link to other attractive states of affairs - that p, because he views what is actually the materialization of p as the materialization of p, i.e. as the materialization of the state of affairs he has attributed to himself the capacity to cause to obtain. The central idea here is that in an intentional basic action the agent's correct view of what is being caused in the action enters into the explanans of this effect. This explains why the context "intentionally brings about that ..." is intensional. It also explains how an intentional action can be done for a reason: the causal power of the state of affairs the agent takes himself to be causing basically (e.g. p) can derive from this state of affairs being linked to others in reason-conditionals. Further, the agent brings about q in an intentional non-basic action, by bringing about p, just in case he knows that p in the circumstances is sufficient for q. In this chapter, the relevant meaning of "basic action" is also explored.

In brief, the outcome of my inquiry into the two issues mentioned as the beginning is this. First, a phenomenal reason of a's for bringing it about that p is a content of his thought expressible in the conditional form "If (only if) I bring about p, q is brought about by me as well", where "q" must represent something wanted by a. Secondly, a's action of causing p and, by this, causing q is governed by this reason just in case: (i) a's thinking that he can cause p (presupposed by the antecedent of the conditional) causes - at least partly due to the conditional relationship between p and q - what a (non-coincidentally) correctly thinks is p, and (ii) a has adequate and true backing for the reason-conditional which is itself true. If it is not already apparent that these explications do not appeal to any irreducible action-theoretical concepts, I hope it will be so by the end of this book. The primitives of the scheme are these: propositional thoughts, truth and reference (of the contents of such thoughts), epistemic relations (between these contents) and causal relations (between these thoughts and e.g. bodily events).

The above outline for the dependence of reasons for action on wants - and features springing from it - is, I suspect, what distinguishes reasons for action from reasons for belief (e.g. the reason for taking a certain pill, that doing so will alleviate one's headache, from the reason for believing or thinking that taking this pill will alleviate one's headache, a reason which might be that the label says so). The possibility of distinguishing between

reasons for action (or practical reasons) and reasons for belief (or theoretical reasons) will simply be assumed from the outset; for although some differences between these types of reasons are pointed out as we go along, there is no room for the systematic exploration of the notion of belief which establishing this conclusion requires.

2. A few remarks on the philosophical method to be employed in this essay may not be out of place. At first glance, this work consists of attempts to set forth truth-conditions for terms of ordinary language like "reason for action", "action", "thinking", "if-then", "wanting", "intentional action" etc. But my real interest is in certain phenomena, namely the phenomena of having a reason for action in one's mind and of acting on it (and phenomena involved in these). This provokes a challenge: why couch an investigation into the nature of certain phenomena in the terms of an inquiry into ordinary language? Could not the classificatory scheme of non-technical discourse be confused and inconsistent or at least crude and indiscriminative, in which case a correct inquiry into it would leave us with a distorted and incomplete picture of the furniture of the world?

My reply to this objection is as follows. There is no way one can avoid presupposing that ordinary language is, to a very large extent, accurate: one simply has to trust that the meaning of certain terms is sufficiently distinct for it to be possible to indicate certain features of the world - otherwise one would be without means of conveying what one's field of study is. Nor is this trust irrational. Ordinary language has evolved gradually during millenia, constantly in close contact with experience, by the method of trial and error.¹ This is surely a pretty good guarantee that, by and large, mistakes have been detected and stamped out. The same facts also provide considerable assurance of the comprehensiveness and subtlety of the distinctions of everyday language: the more protacted the time-span of development and testing, the less the likelihood that something important is being overlooked. This holds especially for discourse about phenomena that occur frequently in everyday experience, such as the phenomena that form the topic of my investigations.

Of course, it is not being maintained that there is anything sacrosanct or infallible about ordinary language. Some phenomena (e.g. mystical experiences) may be so rare and exclusive that no vocabulary designed to cope with them adequately has developed. Furthermore, it happens all the time that scientific research opens up new avenues of experience, the description of which requires the coinage of new terms. Thus, non-technical vocabulary at least stands in need of being supplemented and expanded. It may be said that these termino-

logical inventions concern merely matters of detail and that they leave the basic principles of ordinary language unscathed. But so far as I can see, there is no a priori guarantee that we will not stumble on experiences, the descriptions of which will have profound repercussions on related fundamental and well-entrenched concepts. Still, this is merely a theoretical possibility, and it should not be allowed to upset the hypothesis that in all probability there is a welter of accurate, subtle and far-reaching information about the nature of phenomena to be extracted from ordinary language.

So, the focus of my inquiry is on the world: the study of ordinary parlance is only a heuristic device to gain insight into the workings of the world (in my case, specifically, certain aspects of the workings of mind and body). I scrutinize ordinary language because I believe it contains the most resourceful and at the same time most economical classification of the phenomena I am interested in. This is, to my mind, the rationale behind the injunction that philosophers are to analyze the concepts of common-sense.

I think the methodological considerations I have advanced come pretty close to what J.L. Austin has baptized "linguistic phenomenology" and summarized thus: "we are using a sharpened awareness of words to sharpen our perception of, though not as the final arbiter of, the phenomena" (ibid., p. 182). Austin has been criticized for giving expression here to "a remarkably old-fashioned view of the relation between words and the world".² The view is that there is an experience of reality which is independent of language and against which the adequacy of language can be judged. I fully endorse this Austinian view. I confess this merely in order to show my hand; a satisfactory defence of this position would take us far beyond the limits of the present work (though the issue will be touched upon in 3.4). One consideration (cf. [204], e.g. pp. 64-7) in its favour may, however, be roughly indicated. Learning a language presupposes that the pupil hears what the teacher says, sees what he points to etc.; hence, there must be perceptual awareness that is logically prior to linguistic conceptualization. Moreover, it does not seem plausible to assume that the acquisition of a language radically transforms the pre-linguistic awareness, for if the pupil (not a language user) does not experience roughly the same as the teacher (a language user), it seems impossible that the teacher should be able to teach the pupil to use words with the same meaning as he himself employs. So there must, it appears, be at least a nucleus of experience common to both a language user and a non-language user (though there might be embellishments of this nucleus that only the former experiences).

These methodological reflections have implications for what counts as relevant criticism of the theses put forward in this book. The mere exclamation

"This definition offends my intuition concerning the proper use of the definiendum!" is not enough to move me. Of course, this protest may be a symptom of a failure on my part to capture some significant aspect of the term under scrutiny and so, according to the methodological principles just laid down, of negligence of some essential feature designated by it. But in order to press home any substantial point, it must be shown directly that the latter is indeed the case. In general, I recognize three valid grounds for criticism. (1) Incomprehensiveness: the proposed conceptual system lacks the means of expressing something significant in the relevant field. (2) Incoherence: the system is not internally consistent or consistent with truths that have to be admitted. (3) Inelegance: the system is needlessly complicated or confusedly constructed. If the conceptual apparatus cannot be challenged on any of these accounts, it does not really matter whether it is in agreement with the conceptual structures embedded in ordinary discourse, because the primary objective of providing a system of concepts in terms of which further empirical research into the field can be conducted without confusion, has been attained. That the working hypothesis about the power of ordinary language has been falsified is inconsequential in this connection.

By this method, I hope to cut short disputes about what is a correct account of everyday usage. These may easily become both fruitless and interminable due to one of the following circumstances. (i) People are differently sensitive to linguistic nuances, and there is no agreed measurement of linguistic sensitivity. (ii) Non-technical concepts are, as a rule, vague, and boundary-cases are likely to produce disagreement. (iii) Ordinary language is often, from a theoretical point of view, over-specific (practical purposes having regulated its growth); therefore, it is often necessary, in a systematic study, to suppress certain details - something which may lead to constructions that sound odd. (iv) One may confuse what is merely (pragmatically) misleading and what is downright false. (v) Language is not only employed in literal ways; it is also used in metaphorical, striking etc. fashions, and sometimes it may be hard to say exactly what the literal use is.

It might be objected that some of the points (i)-(v) reveal weaknesses in my philosophical method: someone making use of it may get caught in one of the traps language sets up and be led away from the nature of phenomena. Indeed, it is clear (cf. [176], chap. 6.3) that the pioneers of "linguistic analysis" in the 40's and 50's frequently fell victims to e.g. the confusion listed in (iv). Probably, the selfsame philosophers can be accused of losing their way in the over-specificity of everyday parlance. But this is no objection against their method in itself: no method of discovering truth

is fool-proof, every method can be abused. Moreover, it seems to me that the mistakes mentioned could have been avoided, if the philosophers in question had made clear to themselves that linguistic analysis is only a means to gaining insight into certain aspects of the world.

1. REAL AND PHENOMENAL REASONS FOR ACTION

1. The phrase "to have a reason for acting" is ambiguous in a way that is highly pertinent to my purposes. To have a reason may mean something like to have a reason in one's mind. In this sense, the reasons an agent has are contents of his thoughts or beliefs. But the phrase may also signify that there are certain reasons for an agent to execute some action, whether or not he has thought of them. The reasons an agent has are here not propositions he believes, but propositions that are true. Perhaps the ambiguity of "the reasons an agent has for acting" can be eliminated if one substitutes for it either "an agent's reasons for acting" - which more firmly indicates the first interpretation - or "the reasons there are for an agent to act" - which suggests the second reading. But the original construction is often handy because the two senses of it are not incompatible - obviously, the beliefs of an agent may be true; the phrase may then serve the function of expressing the combination of the two senses.

A reason an agent has in the first sense I shall call a phenomenal reason (for the agent: that something is a phenomenal reason is a relative characterization, but since I always understand the subject to be the agent, I shall not express this explicitly). The reasons a person has in the second sense will be baptized real reasons. It follows from the above that a reason a certain agent has at a certain time can be simultaneously both phenomenal and real, namely if he has a true belief.

The distinction between real and phenomenal reasons is important for the clarification of the two major functions reasons for action fulfil, the explanatory and the evaluative functions. For the explanatory job, only inquiry into the agent's phenomenal reasons at the acting time is relevant. But when the question is evaluation of courses of action - either the evaluation is done by the agent before acting (deliberation) or afterwards by himself or others (review) - the objective is to establish the existence of real reasons, i.e. the truth concerning lines of conduct.

This work will focus on phenomenal reasons, since only reasons actually operative in a person's mind are relevant for what he does for a reason. But the notion of a real reason must not be allowed to disappear entirely out of view: one of the ways of testing the adequacy of an account of phenomenal reasons is by checking how naturally it can be extended to cover real reasons as well. Besides, an appeal to real reasons at certain points may fortify and amplify the account of phenomenal reasons proposed (see (2) in "Conclusions and Limitations").

Kurt Baier has made a tripartition of the contexts where the phrase "reasons for action" is employed into explanatory, deliberative and justificatory contexts.¹ But it is obvious from his discussion that his main purpose is to contrast reasons cited in explanatory contexts, on the one hand, with reasons in deliberative and justificatory contexts, on the other. One point of contrast Baier argues for is the one I have insisted on: that in explanatory contexts the aim is to determine an agent's beliefs about a certain set of possible actions (i.e. his phenomenal reasons), whereas in the other contexts it is to determine the truth about these action possibilities (i.e. ascertain his real reasons).² I have disregarded the differences between deliberation and justification (review) and grouped them together under the head "evaluation". (Note that there are other contexts that belong to the species evaluation, e.g. advisory contexts.)

Obviously, there is a very intimate relationship between real and phenomenal reasons for action. If an agent has certain phenomenal reasons then, of course, these reasons are real from his point of view, i.e. certain propositions are the contents of his beliefs only if he believes that they represent reality. On the other hand, it makes sense to ask what real reasons there are/were for somebody to do something only if this "somebody" is a being capable of acting for a reason, which includes a capacity to have phenomenal reasons. This point may be obscured if one does not keep the idioms mentioned separate from the "reason why" idiom, where the latter merely signifies the availability of any sort of explanation:³ e.g. there is a reason why a flower smells in a certain way, but it is nonsense to say that there is a reason for the flower to emit this smell (provided the latter is meant literally: teleological terms, like "striving" or "purpose", are often applied to vegetation, but I think that these are analogical uses).

As an aside it may be noted that a similar distinction between real and phenomenal reasons can be upheld in the case of reasons for belief, roughly in the form of a distinction between true propositions which really support p and propositions believed by a certain subject to support p.

2. I have said that it is propositions (or what is for me an alternative expression, states of affairs) that figure in the role of reasons for action - more precisely, propositions that are believed (by the agent) or that are true or both. The import of the term "proposition" will not here be characterized any further than has hitherto been implied: that a proposition is something that has truth-value and that can (logically) be the content of thinking-that. (It is noteworthy that not everything that can be the content of thinking-that

has truth-value: someone deciding thinks that he shall do something.) The latter part of the characterization will be sharpened somewhat when (in 3.2.2) I discuss the notion of content of thinking-that. Here I will instead raise the question whether the propositions that can function as reasons for action must necessarily have a specific logical form.

My proposal is that a reason can always be construed as a conditional proposition, a proposition properly expressed by an "if-then" (or an "only if-then") sentence. This conditional will, by me, be so formulated that the antecedent consists of a description of the type of action one gives a reason for, whereas the consequent is a description of a "motivating circumstance". Thus, a reason for the agent a to bring about p⁴ may look like this: (R) (only) if a brings about p, q is (probably, possibly) brought about (by him).

It may be objected that, strictly speaking, it is only what I above called the motivating circumstance that is the reason: what is mentioned in the protasis is what this is a reason for. This objection, however, embodies a subtle fallacy. The antecedent is a description of a possibility - namely, the possibility that a brings about p. What (R) gives a reason for is not this possibility (whatever that may mean), but the actualization of this possibility. Consideration of what happens in a possible world where a brings about p is given as a reason for bringing the actual world into conformity with this possible world. So, the whole of (R) is given as a reason for the actualization of its antecedent.

If a brings about p because of (R), the motivating part of (R) may be said to be (one of) a's purpose(s) or goal(s) in bringing about p.

This thesis concerning the logical form of reasons - that reasons are conditional propositions (that are either believed or true) - is not meant to sum up how "reason for action" is superficially used in ordinary discourse. Sometimes something resembling the motivating circumstance itself is said to be a reason (Anscombe's "forward-looking motives" ([4], pp. 20-1)), but at other times something seemingly unrelated to my standard-formula is named a reason. Consider an example of Anscombe's "backward-looking motives" (*ibid.*): a is said to quit his job for the reason that (as he believes) his boss tried to seduce his (a's!) wife. Appearances notwithstanding, I think my model illuminates even this situation. Why does a react to the audacity of his boss by quitting? Presumably because he views his leaving his job as a way of revenging himself, of showing that the boss cannot treat his employees as he pleases etc., i.e. he thinks that if he quits, he will cause the boss some inconvenience (or he will show the boss that he has a sense of dignity etc.). So, I believe that even in this situation a reason of my model lurks in the

background and is hinted at by the superficially appointed reason.

I find it plausible to assume that whenever we have a reason for an action, a reason of my model can be reconstructed, for there must be some sort of connection between the motivating circumstance and the action supported. My analysis interprets this connection as a hypothetical relationship between a motivating circumstance and a possibility of which the action supported is an actualization. I simply cannot imagine any alternative way of construing this connection,⁵ although I need not claim that my reconstruction is the only one possible, but merely that it is a feasible and fruitful one.

After these preliminaries, the following topics call for treatment. (i) Light must be thrown on the concept of action, since this is necessary in order to understand the more complex notion of a reason-governed action, and since the concept of action enters into the formulation of action-reasons according to my model. (ii) The concept of thought needs to be clarified to some extent if we are to grasp how phenomenal reasons - i.e. conditional propositions that form the content of thought - give rise to actions that qualify as reason-governed. (iii) The conditional form that turns up in the reason must be explicated. We must also consider the question (iv) whether a reason for action must satisfy further requirements beyond those outlined in this chapter. These tasks will occupy me in the following four chapters, respectively.

2. A CONCEPT OF ACTION

I have made use of a concept of action in formulating the conditional proposition that in my opinion constitutes a reason for action ("If the agent brings about ..."). Later, in 5.1, I will assert that whether or not a conditional proposition of the kind mentioned is a reason for a person to act depends on that person's attitudes of wanting and that the relevant wanting is wanting to act. So, in my accounts of reasons for action and of desires a notion of action crops up. Hence, it is incumbent on me to explicate this notion of action without having recourse to either the concept of an action-reason or that of a want, without defining an action as e.g. something resulting in a particular way from phenomenal reasons or wants; otherwise my position will involve a circle or an infinite regress.¹

It might seem that I am overly harsh to myself here. Are not contexts like "to think that ..." and "to want ..." intensional (i.e. co-designative phrases are not substitutable salva veritate in them)? But if so, the objection continues, then even if "to act" is logically equivalent to, say, "to have a want to act (a want) which causes some movement", we cannot without further ado substitute the latter for the former in, say, "to want to act" and so generate an infinite regress. However, this line of defence will not do, because we have to allow for the possibility that an agent acquires knowledge about the conceptual analysis of action. But if we conjugate "a wants to act" and "a knows that 'to act' is logically equivalent to 'to have a want to act (a want) which causes some movement'", then, it seems, one can infer that "a wants to have a want to act (a want) which causes some movement", and the knowledgeable a is caught in an infinite regress.

Action-theorists have differed over the extension of the verb "to act" and related terms: some have held that only speakers of some language can act,² others extend the class of agents to include animals as well,³ while still others let even inanimate things enter the community of agents.⁴ This situation is probably a reflection of there being more than one concept of action in currency. What I would like to stress in this connection is that the last-mentioned group of authors is right about the extension of at least one current concept of action. In non-technical discourse, we do indeed speak e.g. of an acid acting on some metal, of it being an agent. Even in the inanimate world it is possible to draw a distinction between what is active and what is passive, between what acts and what is acted upon. An action-concept of this extension cannot, of course, be dependent on such notions as those of a reason or a want, since these are not applicable to the inanimate world. Therefore,

this action-concept is well suited to my present purposes, and it will be the target at which the following definition aims.

2.1 Outline of a Broad Concept of Action.

Let us start with a simple example of acting in this broad sense: a stone breaks a window. Here the stone is that which acts and the window is that which is being acted on. It is reasonable to seek to analyze this fact in two steps. First, to reformulate it in explicit causal terms as that the stone causes the window to break. It is a familiar idea that in the case of verbs which have both a transitive and an intransitive reading - "bend", "cool", "dissolve", "harden", "melt", "move", "sink", "soften", "warm" etc. - the transitive use could be understood in terms of the verb "cause" and the intransitive, i.e. "a v_t-s b" as "a causes b to v_i" (where "v_t" and "v_i" stand for the transitive and the intransitive reading of a verb, respectively). But, of course, for the translation to work it is not necessary that the intransitive verb, appearing in the analysans, be expressed by the same sound as the transitive verb being analyzed - compare "killing"/"dying" and "raising"/"rising". Secondly, to replace talk of substances as causes with talk of facts as causes, that is, to construe "a stone" as an ellipsis for "some fact concerning a stone" (e.g. the fact that a stone hit the window with a certain impact). Neither of these moves is uncontroversial, and they will accordingly receive some backing as we go along.

If the sentence "The stone breaks the window" is typical for sentences attributing an action to their subjects, the upshot is that such a sentence is equivalent to a sentence saying that some fact concerning the subject caused some other fact. Let "F", "G" and "H" range over all sorts of predicates, not only one-place but two-place etc. ones as well (including their grammatical objects), let "C" symbolize "causes" and "L" logical necessity; then we can define:

(A) "Fa" attributes an action to a iff:

$$\exists(x), \exists(H) \ L(Fa \leftrightarrow \exists(G) \ (Ga \ \& \ Hx \ \& \ C(Ga, Hx))).$$

To illustrate this definition: suppose "Fa" is "a moves a finger of his". This is - according to the analysis of transitive verbs advanced above - equivalent to "a causes a finger of his to move", which in its turn - granted the principle that a substance is a cause iff its instantiation of some property is a cause - yields "There is some property such that a has it and this instantiation causes a finger of his to move". Here the finger = x, H = move; and G might be some mental predicate like feeling an itch somewhere within the vicinity of the finger. So, we see that "a moves a finger of his" satisfies

(A)'s requirement for ascribing an action to a.

I shall, appropriating a term originally introduced by G.H. von Wright,⁵ baptize the effect-fact, Hx, the (intrinsic) result of Fa. Observe that, though Hx is an effect, it is not an effect of that of which it is the (intrinsic) result, namely Fa, but of Ga; Fa entails Hx. The definite article - "the result" - is used to signalize that what I have in mind is the most comprehensive fact the causation of which is inferable from Fa. Thus, if F is to move a finger swiftly, the result is that the finger moves swiftly, not merely that it moves, although the latter is a result.

My employment of the term "result" diverges from von Wright's. Von Wright draws ([216], pp. 41-2) a distinction between acts - such as closing a window, lighting a cigarette etc. - and activities - such as running, smoking etc. He then declares:

"Activity is not related to changes and to states of affairs in the same manner in which acts are related to their results" (p. 41).⁶

What seems characteristic of von Wright's acts is that there is some state of affairs the materialization of which necessarily terminates the act: thus, when the window is closed or the cigarette lit, the act of closing the window or lighting the cigarette is necessarily completed. But no such terminating state of affairs is entailed by "smoking" or "running". The quotation indicates that von Wright reserves the term "result" for such terminating states of affairs - or changes ending in them - and, consequently, activities will have no results.

For my own part, I want to insist that activities as well have results, since in them too something is necessarily caused: in running, legs are moved, and in smoking, arms are moved and muscles in the chest contracted etc. Von Wright would be hard put to defend the thesis that only in acts is something caused or brought about (if he should be so inclined), for the distinction between acts and activities only exists at the level of description, as a distinction between different ways of describing one and the same thing. It is necessarily true that if a engages in some activity, say, running, he performs some act, say, running a certain distance or for a certain time. (The converse holds as well: if a performs some act, he must engage in some activity at least for a short span of time.) This implication guarantees that if something is brought about in an act, something is brought about in an activity as well. So, although I by no means deny the possibility of drawing a distinction between acts and activities,⁷ I wish to stress a feature they share: in both of them, something is necessarily caused, i.e. they have results in my sense.

It must be conceded, though, that in the case of some activity-verbs where the transitivity/intransitivity distinction does not apply unproblematically - e.g. "climb", "eat", "hum", "kick", "laugh", "spit", "swallow" - it is a considerably subtler task to extract the fact being caused than it is in the case of the verbs earlier used to illustrate (A). But I cannot see why the effect-fact should be mechanically retrievable (contrast [61], p. 119).

It would, I think, cast light on the purport of (A) if I in more detail indicated how the line between bodily actions and bodily processes or reactions like perspiring, bleeding, trembling, growing, blushing, sneezing, vomiting, dying, fainting etc. is to be drawn. Irving Thalberg suggests that our refusal to classify the latter as actions is based on a belief in certain empirical truths about the limits of human powers:

"All that disqualifies bodily process and reaction verbs from being action verbs, and bars the episodes in question from being things we do, is our belief about human powers. We believe, on good evidence, that people do not have the control over their bodily processes and their reactions which they have over their walking and stealing" ([200], p. 62).

If people acquired the same control over their perspiring, bleeding etc. as they presently exercise over their finger-movements, then, Thalberg argues, "perspire" and the rest would "migrate" (as he expresses it) into the category of action-verbs. But this is a confusion. The fact that I have (direct) control over my finger-movements_i does not turn "move_i" into a verb of action; all that follows is that there is room for "move" being used in a transitive manner as well - i.e. as being equivalent to "(directly) cause to move_i" - and this is indeed a verb of action. Similarly, if people established the same rule over their perspiring etc., "perspire" might well acquire a transitive use - being equivalent to "(directly) cause one's body to perspire_i" -, and this transitive verb would qualify as an action-verb. So, the extension of direct control might effect the advent of a new sense of an old sound, but not the migration of the sound with its old sense from the category of bodily process verbs to action-verbs. It is inconceivable that a type of bodily process or reaction should become a type of action: that would mean that a type of non-causal fact would be transformed into a type of causal fact.

Thalberg's mistake put in a nutshell seems to be this: he confuses the result of an action with the action itself, i.e. the fact that p with the fact that something causes that p. (This mistake is more easily committed than it may seem, since a transitive and an intransitive verb are often expressed by the same sound.) Sentences like "a perspires", "a bleeds" etc. do not ascribe actions to a because they do not entail that he caused anything; but this is quite consistent with their reporting results of his

actions, e.g. perspiring might be brought about by exercising.

2.2 Comparison of Some Theories of Action.

At this point, it might clarify matters to compare and contrast some theories of action.

(1) The theory embodied in (A); the fact that a acted is here identical to a causal fact that something's becoming the case about a caused something else to become the case. The fact that an action occurred is conceived as a relational fact, involving two facts standing to each other in the relation of causation.⁸

(2) The theory that identifies the fact a acted with the fact that some result occurred as due to a certain cause within a. This type of theory holds that "a moved a finger" is equivalent to "An effect of some fact about a, namely (the effect) that a's finger moved, began to obtain". We must not confuse this theory with that under (1): that would be to confuse a sentence of the structure "An effect of p, namely q, began to obtain" with one of the structure "That p (began to obtain) caused q (to begin to obtain)". Only the latter states that p came to obtain and that this caused the materialization of q; the former merely presupposes these facts by referring to q by means of a certain relational description ("an effect of p"): the only fact stated here is q. This means that, according to the theory under this head, e.g. "a moves a finger of his" states the same fact as "A finger of a's moves"; the difference is only that in the former sentence the fact of the finger's moving is also identified in a relational way as an effect of a certain cause. Monroe C. Beardsley seems to be a proponent of this theory when he affirms: "Z's bodily deeds, then, are those of Z's bodily events whose proximate causes are Z's muscular behavior".⁹ Here the class of a person's bodily deeds seems to be identified with a subclass of bodily events like e.g. movements; of his body, to wit, the subclass of such events that has a specific cause. A.I. Melden seems to voice ([138], e.g. pp. 73-4, 185ff) a non-causal version of the same type of theory.

When we turn to actions whose descriptions refer to changes outside the agent's body which are caused by movements of the body, e.g. acts of killing, typing and melting chocolate, the theory under consideration can take two forms. (a) One can go on identifying the fact that e.g. a killed b with the fact of the materialization of the result, i.e. that b died, described as the effect of a certain type of cause. This appears to be the course Beardsley takes ([26], pp. 267, 270). (b) But one can also let the descriptum of the action-description remain constant at the level of bodily movement and conclude that what happens when an action-description refers to events outside

the agent's body is merely that a bodily movement is redescribed in terms of its effects. According to this proposal, "a killed b" would be equivalent to "A bodily movement of a's, which is the effect of certain facts about a and the cause of b's death, occurred". This redescribing of a bodily movement in terms of its effects provides the background for Davidson's contention that "(w)e never do more than move our bodies".¹⁰

(3) The type of theory that identifies the fact that an action occurred with the fact that a certain cause occurred. Here "a moved a finger" will be equated with "The cause of the fact that a's finger moved, namely a certain type of fact about a, began to obtain". H.A. Prichard's view [162] that our actions are identical to our willings is of this type (though he deals with a more restricted concept of action). This type of theory has quite recently come into vogue; it has been revived by such authors as Bruce Aune (with some reservations, [13], chap. 1), Lawrence H. Davis ([72], p. 41) and Jennifer Hornsby ([104], chaps. 1 & 2).

The question now to be asked is: why is (1) preferable to (2) and (3)? (1) is superior to (2) and (3) in that it allows of the most natural locating of actions in time. Suppose that a stabs b at t_1 and that, as a consequence of this, b dies some time later, at t_2 . When did a kill b? According to (2a), a killed b at t_2 or, more precisely, the killing occurred in its entirety after the stabbing, since it is to be identified with b's dying. But this is surely counter-intuitive, for although the killing may be completed after the stabbing, it is not in its entirety to be located after it, since they appear to commence simultaneously. (2b) also implies an implausible locating of the killing, because it entails that b is killed before he dies - remember that the killing is here identified with the stabbing movements which occur at t_1 . Finally, (3) has repugnant implications for the dating both of the killing and the stabbing. Since the cause of the occurrence of the stabbing movements is present before these movements take place, say at t_0 , it will follow that a stabbed b before the stabbing movements happened at t_1 and that he killed b before b died at t_2 . In sum, these theories are implausible because we are intuitively inclined to judge (i) that an action includes what I have called its result and, hence, that it is not completed until the result has materialized (this rules out (2b) and (3)) and (ii) that an action, at least sometimes (e.g. when the result is a change outside the agent's body), starts before its result materializes (which precludes (2a) as well). The superiority of (1) consists in its taking care of both (i) and (ii): it absorbs (i) because an action is here a causing, and the effect is a component of a causing, and it absorbs (ii) because a causing commences when the cause occurs. (See

further remarks on my use of "cause" in 2.3 under (1).) A causing is an event which commences with the occurrence of the cause and ends with the occurrence of the effect. (I will in 2.4 deal with the objection that (1) as well has a counter-intuitive corollary.)

Let me review two further theories of action with which (1) or (A) is incompatible.

(4) The theory that rejects the transcription of "a causes ..." (where "a" is a constant for human subjects) as "some fact (or event) concerning a causes ..." and claims that agent-causation is an irreducible form of causation. It has been maintained by Roderick Chisholm and Richard Taylor¹¹ that "immanent" causation, or causation by a human agent, constitutes a peculiar variety of causation distinct from "transeunt" causation, causation by facts or events. Chisholm's main reason for introducing the concept of immanent causation appears to be his belief that responsibility is at odds with both indeterminism and "transeunt" causal determinism. I cannot here query this belief, but I would like briefly to challenge the contention that immanent causation presents an intelligible third alternative in between "transeunt" determinism and indeterminism.

Suppose that a moved a finger of his at t; then if the cause of the movement is strictly a, in contradistinction to something that became a fact about a at t, one wonders why the movement of the finger did not occur earlier - after all, a has presumably existed for a long time before t. Is the reason that a at t (or immediately before) began to exemplify the properties P₁, ... P_n, that the instantiation of these properties is a necessary condition for the finger-movement? Perhaps. If so, the instantiation of these properties is perhaps a sufficient condition as well: possibly, it is a universal law that (x)(t) if x instantiates P₁, ... P_n at t, x's finger moves at t. But then it seems that immanent causation after all collapses into "transeunt" causation. So imagine that we cannot get beyond a generalization that holds only for some x and t - but is this not merely what a transeunt indeterminist could contend? So it seems hard to escape the conclusion that there is nothing intelligible between "transeunt" determinism and indeterminism.

(5) The theory that the notion of action is prior to that of causation. My account rests on the thesis that the concept of action is secondary to that of causality: the agency of a substance is reduced to the causal efficacy of some fact concerning it. Von Wright reverses this relation of dependency and argues instead that "the idea of a causal or nomic relationship can be said to depend on the concept of action".¹² The reason for this view lies in the role he assigns to action or experiment in the verification of causal judgements.¹³

Von Wright asserts that what distinguishes a causal or nomic relation from a mere regularity is the element of counterfactuality the former contains: "At t, p causes q" entails "If p had not obtained at t, q would not have obtained either at t". According to von Wright, the nearest we can come to the unattainable ideal of a conclusive verification of this counterfactual is by doing p on an occasion when neither p nor q is present and observe that q then follows.

I would like to make three objections to this argument of von Wright's. The first and most radical objection - which will be developed in 5.4.4 - is that conclusive verification of singular causal judgements is possible even where we are completely passive, viz. when we are aware of the impact of an external object on our body. This, if correct, is fatal to any attempt to tie the notion of causality to the idea of human interference. Secondly, experimentation seems to have at most a value as a time-saving device: it seems that we could in theory come as close to verification of counterfactuals simply by waiting until p materializes independently of us. Neither of these methods - manipulation or patient waiting - could decisively exclude rival hypotheses - e.g. the presence of a factor of which p and q are joint effects - but the latter method could make these hypotheses just as improbable as the former, given that we are in no hurry. Thirdly, even if manipulation were indispensable in principle in the procedure of verifying causal counterfactuals, nothing about the conceptual relationships between action and causation immediately follows from this. A "verificationist" theory of concept-formation is needed as an extra premise, and such a theory has not been stated clearly, let alone defended (as von Wright seems to concede, ([217], p. 195)). So much for undermining the support von Wright provides for his thesis that the notion of action is prior to that of causation. Let me now argue that the thesis itself has awkward consequences.

In order to avoid circularity, von Wright must deny that causation is involved in the notion of action appealed to in elucidating the concept of causality. I think that von Wright's view is that the concept of causation rests on the notion of a basic action, "on the idea that man can do things, as distinct from bringing them about, though a direct interference with the course of events (nature)" ([217], p. 66). But von Wright is not as univocal on this point as one might wish - for example, his example of turning a sheet of paper ([218], p. 43) might easily mislead a reader into supposing that he holds the absurd view that causality does not even enter into non-basic actions. (For a discussion of the distinction between basic and non-basic actions, see 2.4.) But this would not be consistent with other parts of von Wright's doctrine, e.g. his distinction between results and consequences of

actions (see e.g. [217], pp. 66-9). Thus, all von Wright should maintain is, in the words of Frederick Stoutland, that "the performance of a basic act never presupposes that anything has been caused to happen" ([197], p. 310).

According to a causal analysis of action, as I shall unfold it, the proposition that a performed a basic action at t, say, moved a finger, entails the counterfactual that if a had not moved the finger at t, the result of the action, that the finger moved, would not have materialized either (everything else being the same). On my view, this counterfactual is, of course, at bottom causal: it says that had the result not been caused by a fact concerning a, it would not have materialized at all at t. Now von Wright disputes that the statement that a basic action has been executed entails a counterfactual of this type which connects the occurrence of the action with that of its result: he argues ([218], pp. 131-2) that it might be true that an agent performs the basic action of getting up from a chair, although there is e.g. a hidden mechanism at work which would have lifted him from the chair in any case. But still he does not reject the view that there is an element of counterfactuality in the concept of a basic action; he modifies the thesis as follows:

"The element of counterfactuality consists in that the agent confidently thinks that certain changes will not occur unless he acts" ([217], p. 199).

"What is excluded /by the notion of myself acting/, is that at one and the same time I raise my arm and observe the operation of the cause ... When I observe, I let things happen. When I act, I make them happen. It is a contradiction in terms both to let and to make the same thing happen on the same occasion" ([217], p. 130; cf. [218], pp. 122-33).

Thus, according to von Wright, the counterfactual element in a basic action consists not in the truth of a counterfactual asserting that the result would not have obtained, had the agent not acted, but in the agent's conviction (at the acting time) that this counterfactual is true (cf. [197], pp. 312ff).

I think that there is a great deal of truth in von Wright's positive view of the counterfactual element in action (given that the concept of action is a restricted one which involves reference to intention),¹⁴ though I will later (in 2.3) argue against his claim that the counterfactual need not be true. The question now to be raised is, however, how von Wright construes the counterfactual which forms the content of the agent's belief. He is aware ([217], p. 72) that if this counterfactual is taken to be causal, his position will be circular.

In von Wright's view, to say that an (intentional) action has taken place is to say that the appropriate propositional attitudes (intentions, beliefs) were present, but not that they caused the result of the action, the behaviour. Instead it is to say that they created a setting that made a certain inter-

pretation of the behaviour possible:

"Behavior gets its intentional character from being seen by the agent himself or an outside observer in a wider perspective, from being set in a context of aims and cognitions" ([217], p. 115).

If the propositional attitudes that turn a piece of behaviour into an (intentional) action in no way have to produce it for this redescription into action-language to be legitimate, then it is, at least prima facie, hard to see what could be meant by counterfactuals like "If a had not wanted (intended or otherwise been motivated) to raise his arm at t, his arm would not have risen at t". Yet it seems certain that a's action (in an intention-laden sense) of raising his arm must involve belief in this type of counterfactual if it involves belief in any type - surely, the agent must believe that his forming certain wants or intentions makes a difference to the movements of his body. Something like this must lie behind the conditional "If the agent had not raised his arm, the arm would not have risen", belief in which von Wright requires for an instance of the action of arm-raising. The latter counterfactual could surely not be upheld merely on the ground that the agent's action of raising his arm is construed as the same event as his arm going up, in which case the counterfactual comes out as the platitudinous "If a's arm going up - which took place in an intentional context - had not occurred at t, a's arm going up would not have occurred at t". The counterfactual, belief in which is required, ought obviously to be synthetic, having to do with the origin of the result.

Von Wright has some things to say about his counterfactual that sound strange to my ears. He characterizes confidence in it as "a confidence in the absence of causes which would disrupt the continuation of a state of affairs which I have come to regard as normal" ([218], p. 43), and not as a confidence in the continuation of this state of affairs as due to some cause which my interference counteracts. But suppose a limb of mine moves due to an external cause and that my flexing some muscles puts an end to the movement; then it seems a blatant falsehood that, in acting, I must regard the perpetuation of the movement as of this type: a "causeless continuation of certain normal states of affairs" ([218], p. 43). Surely, I definitely do not consider it to be causeless. This attempt to distinguish the counterfactual involved in action from the causal one is therefore a failure.

I think we are forced to conclude that we need to appeal to the notion of causation in explicating the element of counterfactuality involved in the concept of action (whether this element consists in the agent's belief in, or the truth of, the counterfactual that had the agent not acted the result would

not have obtained). I have also endeavoured to show that von Wright has provided no compelling reason for abstaining from this appeal.

2.3 Further Specification of the Broad Concept of Action.

After this exposition of my theory of action in broad outline and its defence against rival theories, let me now try to make the import of (A) a little more precise. My remarks can be grouped around three aspects of (A): (i) the notion of causality employed, (ii) the range of the predicate-constant "G" and (iii) the sense in which "Ga" must be about a.

(1) I shall in this work make the following assumption about the ontology of the relata of the causal relation: in a particular instance of causation, particular facts, rather than individual events, are united by the causal relation. A particular fact, as I understand it here, is an instantiation (exemplification) of a property in a certain thing or substance (or in a certain place) at a certain time. An individual event is roughly an instantiation of an indefinite set of properties in a certain spatio-temporal zone or in a substance at a certain time, when the properties of the set are such that their coincidental instantiation is a matter of some necessity (logical or nomological). The property the exemplification of which constitutes the fact is revealed by the predicative part of the "that"-clause that conveys the fact; but only a fraction of the properties whose instantiations make up an event are disclosed by the gerundive construction (i.e. a construction of the form "x's F-ing at t") used to refer to an event. These rough characterizations of the notions of a fact and an event should be enough to enable one to gauge what the advantage is of casting facts in the roles of cause and effect: in this way, citations of cause and effect become more precise, since it can be exactly displayed which property the exemplification of which causes the instantiation of which other property. Thus, my claim is not that it is absurd or illicit to let events function as cause and effect (it is even grammatically permissible to put substances in these roles), but that it is a more diffuse way of specifying them (unless one talks about events-under-descriptions which is in effect equivalent to fact-talk). (In order to avoid unnecessarily cumbersome formulations this norm for phrasing causal assertions will not be adhered to without exception.)

However, it has been argued, e.g. by Davidson ([62], pp. 85-6), that construing causes and effects as facts has the paradoxical consequence that the context "... causes ..." becomes truth-functional, given (a) that the context is referentially transparent (allows of substitution of co-referential phrases) and (b) that substitution of logical equivalences in the context is

permitted. I think that the root of the untoward consequence is in assumption (b). As I have outlined the concept of a fact, property-identity is necessary for fact-identity. But it seems to me that the logical equivalence between two sentences in no way guarantees that the predicate-parts of them express the same property. For instance, "a is F" is logically equivalent to "a is F and G or -G", yet I doubt that "is F" and "is F and G or -G" express the same property; "The F and G is H" is logically equivalent to "The F and G is G and H", although it is false that "is H" and "is G and H" stand for the same property. So. I am inclined to extricate myself from the difficulties mentioned by denying (b); but of course much more needs to be said about this issue than there is room for in this work.

A necessary requirement for p being the cause of q at t is that p is at least what J.L. Mackie ([128], p. 62) has called an inus-condition for q at t, i.e. p is at least an insufficient but necessary part of a condition that is unnecessary but sufficient for q at t. (If p itself is a sufficient condition - let alone the only sufficient one, i.e. necessary as well - for q, the requirement is satisfied as well.) That p is a necessary part of such a condition does not mean that p is irreplaceable - i.e. that nothing else, if it were the case at t instead of p, would make the condition sufficient - but merely that p is non-redundant, i.e. that without it, everything else being what it is at t, the condition would not be sufficient. But I would like to strengthen this requirement in two respects.

First, let me add that p must not only be a non-redundant part of a sufficient condition but of every sufficient condition for q that is present at t. (For the sake of brevity, I will assume that there is just one such condition and speak of it as the sufficient condition.) This addition is designed to be an insurance against over-determination - especially in the shape of there being present another sufficient condition for the intrinsic result, a condition which has nothing to do with the agent. In opposition to von Wright (see above p. 20), I do not wish to say e.g. that a moved his finger at t if it is true that if he had not moved his finger at t, his finger would still (everything else being the same) have moved at t. Imagine that a and b simultaneously fire at c, that their bullets at the same time pierce his heart and that each bullet is sufficient and just as quick-working as the other - who caused c's death? My answer is: a and b caused it together, but that you cannot infer from this e.g. that a caused c's death.¹⁵ There is nothing strange about the invalidity of this inference. Suppose that neither shot had been sufficient without the other, but that they were jointly sufficient (in the circumstances); then it is true that a and b jointly caused c's death, but it does not follow

that a did it or that b did. The case of over-determination is merely an extreme instance of this situation.

Secondly, p must be a state of affairs such that when it begins to obtain, it completes the sufficient condition for q at t; all the other non-redundant components of the sufficient condition must obtain beforehand. In other words, p must be what triggers off q at t. I am not putting this forward as a general restriction on the use of the word "cause" but rather as a condition that must be satisfied, if the subject of p is to be considered an agent of q. We might well count a so-called standing condition as a cause and say e.g. "What caused a to drop the jar was the fact that the handle was greasy"; but we would not on this ground say that the jar was the agent of the fact that a dropped it. From this point of view, this requirement - that the causing fact must be a fact of change, that it must consist in a property beginning to be instantiated - can also be viewed as a restriction on the predicates "G" is intended to range over, a restriction which excludes predicates like "is red-haired", "is called Algernon" etc., predicates the exemplification of which constitutes static facts.¹⁶

Is the causal relation, as I have characterized it (i.e. as entailing that a cause is a non-redundant part of and completes the sufficient condition), transitive? A moment of reflection will reveal that it is not, for a situation of the following type is imaginable. p is materialized at t₁, and this causes q to obtain at t₃. When q comes to obtain at t₃, this completes a sufficient condition for r to become a fact at t₄, but only because s - independently of p - has come to obtain at t₂. In this situation, p causes q, and q causes r, but p does not cause r, since the sufficient condition for r is not completed until t₂ when s has materialized.

This result is of some importance because it has been objected ([80], pp. 130-1; [202], pp. 128-9) against a causal analysis of transitive verbs (in the grammatical sense) that it cannot be right because of the transitivity of the causal relation, e.g. that "kill" cannot mean "cause to die", because if a causes b to kill (ex hypothesi, to cause the death of) c, then it does not follow that a has killed c, though he (it is alleged) has caused c's death, in virtue of the transitivity of "causes". The weakness of this argument is that it takes for granted that the causal relation (as it figures in the analysis of agency) is transitive. At this juncture, the counter-objection might be: the argument could get along with the weaker assumption that the causal relation is not intransitive, i.e. that the transitive inference is sometimes valid - and, specifically, that it often holds good in situations where a causes b to cause the death of c, whereas it is never the case that

a kills c in these circumstances. But I want to deny both of these claims. First, it seems to me in general strained to say that a caused the death of c: when b does not automatically give in to a's influence, it is doubtful whether the transitivity-inference goes through. Secondly, I can afford to admit that, though it is often to strain matters to say this, it is perhaps not always inappropriate, for we sometimes also say that the person in authority killed the victim although, strictly speaking, he only gave the order and his underlings executed it (i.e. caused someone to cause the death), e.g. "Hitler killed six million Jews".¹⁷

One final comment before the topic of causality is left. Needless to say, one must distinguish between "not causing p" and "causing $\neg p$ ": the former does not describe an action, whereas the latter describes the action of preventing p from materializing (when it would have, had one not interfered)¹⁸ or that of eliminating or abolishing p when it already obtains.

(2) I place no restriction on the material content of the predicates over which "G" ranges; they might be e.g. both mental and physical. As we have seen (p. 16), Beardsley holds that a bodily event is a bodily deed only if its proximate cause is muscular. But this cannot be necessary, for why should one not be capable of moving one muscle (as an action or a deed) without first causing another muscle to move?

In order to illustrate the breadth of the action-concept defined, let me adjust an example of Davidson's ([66], p. 17): suppose a wave upsets the balance of an officer so that he falls on the button which fires a torpedo which sinks a ship; then the officer may have performed the action of sinking the ship: the fact concerning him that he fell on the button will do.

To someone accustomed to narrower concepts of action - like e.g. the Davidsonian one - this concept may seem excessively wide; but the concept has some attractive consequences. For instance, in contradistinction to narrower concepts, it allows one to make sense of modifiers like "involuntarily", "on reflex", "habitually", "mechanically", not only ones like "intentionally", "at will", "deliberately" - that is, to make sense of the former as modifying the description of the same type of phenomena as the latter, namely actions. It is natural to regard also the first set of modifiers as operating on a very broad action-concept and adding qualifications to it. If a definition like (A) is accepted, it becomes very easy to reconstruct this procedure of qualification: it is achieved by imposing restrictions on the causing fact or, more accurately, on the material content of the predicates "G" accepts as values. This is the form my definition of acting at will or intentionally, in chap. 6 will assume.

But I would like to put one further formal constraint on the domain of values of "G": only predicates in the active voice are accepted as values. A sentence like "b is thought of by a" should be rephrased as "a thinks of b" before it is formalized. The formalization then becomes "Ga" ("G" = "thinks of b"), not "Gb" ("G" = "is thought of by a"). So, if what makes these sentences true is the cause of anything, then a, not b, is the agent according to (A).

(3) That a must be the subject of the causing fact should not be interpreted as being incompatible with it being true that, strictly speaking, only part of a is the subject. If a's arm falls on a bug and crushes it, a might still be said to have crushed the bug, although the causing fact, strictly speaking, involved only a part of a, his arm. The reason is that when a's arm moves, a can also be said to move_i, that is, the predicate "move" can here be loosely applied to the whole of which the moving subject (in the strict sense) is a part. I shall not here embark on any investigation into the conditions that part-whole relationships must satisfy for a predicate being also ascribable to the whole when it is strictly ascribable to a part of it. Suffice it to say, that the predicative part of the cause-describing sentence must be true of a, at least in a loose sense.

So much for the elucidation of the broad concept of action. The details of this definition do not matter much; they can be settled pretty arbitrarily. The definition is important, however, as an indication of how a concept of action can be defined without appeal to such notions as those of intention, reason for action etc. A general feature of the concept is also noteworthy because it is preserved in the notion of an intentional action: that to state that someone acts is to state a causal fact, which incorporates a causing fact as well as a caused fact (the result of the action). This is important since it stresses the need for a distinction between the causes of an action and the causes of the result of an action. I shall later argue that it is part of the concept of an intentional action that its result has a certain type of cause - that is what I mean by giving a causal analysis of intentional action - but it is not part of this concept that the action itself has any particular type of cause.

2.4 Some Distinctions in the Domain of Action.

One much discussed distinction in the theory of action is that between basic and non-basic action,¹⁹ that is, between actions done directly and actions done in or by doing other types of action. It is often held - e.g. by writers like Davidson [66], Arthur Danto²⁰ and Raimo Tuomela ([205], pp. 292-4) - that all we ever do in basic actions is move our bodies. As concerns

non-basic actions, the most important type is presumably causally non-basic actions. a brings about q in a causally non-basic action by bringing about p or, as Alvin Goldman ([88], p. 23) expresses it, his bringing about p causally generates his bringing about q just in case a brings about p and his doing so - or rather the fact that p materializes - causes q to materialize (in such a way that transitivity holds).

However, at the level at which we are now operating - the ontological as opposed to what in chap. 6 will be designated the teleological - I do not think that it is feasible to distinguish between basic and non-basic actions in a non-relative way. (I am not suggesting that the writers just referred to operate on the ontological level.) Consider the definition of causal generation: is it not true that when a person causes his body to move, he does this by causing muscular contractions which cause the body to move? So, it appears that moving one's body is not an ontologically basic action, but (from an ontological point of view) a causally non-basic action. But the same reasoning can be applied to the action of causing one's muscles to contract - is this not caused by e.g. causing neurons to fire? It seems difficult to decide what is the first effect, the result of the ontologically basic action. But supposing this could be decided, would the class thus singled out be of any peculiar theoretical significance? Not so far as I can see.

For these reasons I shall be content to apply the notion of basicity only in a relative or comparative way at this ontological level: causing one's muscles to contract is ontologically more basic than causing one's body to move, which in its turn is ontologically more basic than casting a shadow on a wall. As a substitute for an absolute basic/non-basic dichotomy and a classification of different types of non-basic actions, it will be convenient to have a rough-and-ready classification of actions on the basis of what their results contain. If an action (a slice of a causal process) consists in nothing but (what is involved in) a causing of a macroscopic fact concerning the agent's body, I shall speak of a bodily action; if it also includes an effect produced in the external world (including the agent's body being acted upon as if it were a foreign object), we have a manipulative action; and for an action which encompasses some circumstance outside the agent's body, whether natural or conventional, but no effect concerning it, I will coin the expression "circumstantial action". Stretching one's arm out until it touches a table is a circumstantial action, since its description entails that a table exists in the vicinity, but nothing about any effect concerning it. Breaking a promise is another example, of the conventional variety. The class of circumstantial actions includes those actions which Goldman speaks of as

conventionally generated and some of those he considers to be generated by simple and augmentation generation ([88], pp. 20-30). The parallel is even closer to Thalberg's "conventional" and "relational" non-basic actions ([201], pp. 89-92).

I would also like to draw a distinction which cuts right across the set of manipulative actions, namely a distinction between essential and consequential actions. Sometimes a line is drawn between an action itself and its consequences, while at other times the action is described in such a way that what could be regarded as its consequences are incorporated in it. For example, b's death might be viewed as a consequence of a's action, but a's action might also be described as a killing of b. My suggestion is that in the former case we employ the concept of an essential action and in the latter case that of a consequential action. An essential action is an action whose result is such that if it comes to obtain at t, something must be the case about the agent at the very same time or (if causes necessarily precede their effects) just before t, something which causes the result. In the case of a consequential action, the result must, of course, also be caused by some fact pertaining to the agent - otherwise he would not, according to (A), be the agent of it -, but this fact may belong to the distant past when the result comes to obtain. Bodily actions are essential and so, normally, are a lot of the manipulative actions, e.g. writing with a pen, lifting something, bending something, whereas killing somebody, blowing up a bridge etc. often are consequential actions; the former are actions which cease when the causes within the agent cease, the latter frequently go on after that point.

This is a expedient distinction for me because it provides me with the means of extricating my analysis of action from a paradox it might otherwise be saddled with. Since I affirm that an action is not completed until its result comes to obtain, I am forced to accept that a person's action might not be completed until long after he has died, for the result might not materialize until then: e.g. in the play, Laertes does not kill Hamlet before his own death. This, it might be urged, is paradoxical.²¹ But the air of paradoxicality vanishes once the distinction between essential and consequential actions is recalled. Of course, a person cannot perform an essential action after his death, since at that time there can be no facts constituting his personhood (but his body may still act, e.g. cause someone fear). There is no reason, however, why a consequential action - such as e.g. a killing - should not extend itself in time beyond the agent's death: it is part of the definition of a consequential act that its result may obtain at a time when nothing causally relevant to the result need be true of the agent.

These considerations might be used to undermine another argument of Jerry Fodor's against a causal analysis of transitive verbs. Fodor's argument is perhaps best formulated as that which makes a sentence like "a caused b to die on Sunday by stabbing him on Saturday" true, is not the same as what makes a sentence like "a killed b on Sunday by stabbing him on Saturday" true.²² The contention that the latter is not true when the death occurs on Sunday due to the stab on Saturday can be queried: if the killing did not take place (in part) on Sunday, when did it take place? Surely not entirely on Saturday - one day before b's death. However, it seems that there is something unnatural also in saying that the temporal location of the killing included the Sunday as well, for it is conceivable that a was passive on that day. I can afford to concede this, and explain it as follows: "kill" hovers between "cause to die" and "do an essential action (i.e. cause something subject to the temporal constraint just laid down) and this action causes a death". (These are not always indistinguishable, since "cause" is not transitive.) If "kill" carries the latter sense, it is not correct to say that the killing occurs in its entirety on Sunday, since the essential action is completed on Saturday. On the other hand, it is not correct either to say that the killing takes place entirely within the space of Saturday, for the essential action takes effect on Sunday. Assuming that "kill" is analyzable in this conjunctive fashion would explain the bafflement the dating of a killing is bound to evoke. I could afford to accept this latter reading of "kill", since it does not interfere with my vital point that an action is a causing.

Slightly resembling the distinction between basic and non-basic actions is the distinction between simple and complex actions.²³ An action is complex just when it is composed of at least two generationally independent actions, and it is simple just when it is not complex. When somebody e.g. clenches his fist, he is bound to perform a series of actions, which do not bear the relationship of generation (in Goldman's sense) to each other, such as bending the index-finger, the middle-finger etc.; therefore this action is complex. Like the notion of basicity, that of simplicity can only be applied in a relative or comparative way at the ontological level, for every action has some spatio-temporal extension, and this is enough to guarantee that it can be divided into more minute actions. But we can speak of one action being ontologically simpler than another, e.g. of bending a finger being simpler than clenching a fist.

3. THE CONCEPT OF THINKING

I will of course not attempt to cover every controversial aspect of the intricate concept of thinking in this chapter. Rather, I will concentrate on those features that are especially urgent to my purpose. My interest in the topic of thought stems from the fact that I plan to analyse such notions as acting for a (phenomenal) reason or acting at will as thoughts of a certain form causing certain changes. It is thus features of thinking in different ways pertinent to this thesis that I will try to highlight. In 3.1 a major purpose is to show that thinking does not involve any mental act(ion)s of "assent", which are to be analysed partly in such practical notions as wanting, deciding etc. Hence, my thesis about acting at will runs no risk of being circular on this score. In 3.3 I aim to make my reading of "thoughts cause behaviour" clearer by contrasting it to an alternative reading and criticizing the latter. The section in between, 3.2, is dedicated to the tasks of further elucidating my notion of a proposition as something "thinkable" and of expounding my views on the intensionality of the context "thinks that ...". The latter is necessary if one is to gain insight into exactly what is done at will. Lastly, in 3.4, I discuss the question whether it is possible to think in any other medium than that of conventional languages. I argue for an positive answer and re-introduce the unpopular view that thinking can be carried out in a medium that at least partly consists of images. The phenomenological credibility of my theory of acting at will is, I think, decreased if thinking must be of a wholly verbal character.

3.1 Thought and Belief.

1. I am here concerned with the sense of constructions like "a thinks that p" when they record an episode (and not a disposition), as e.g. "For a moment I thought I would lose my balance".¹ The first thesis I shall attempt to make plausible is:

(T1) "a thinks that p" entails "a believes that p".

As (T1) makes clear, I do not equate thinking that p with what has traditionally been called "entertaining the proposition p". H.H. Price regards the latter as being a synonym to the non-technical expression "thinking of the proposition that p" ([161], pp. 190, 203). Clearly, merely thinking of the proposition p does not involve believing or endorsing it. Hence, thinking that p is not to be identified with thinking of the proposition p.

Less obvious is that the former does not even include the latter. When I think that Seth is fat, I think of Seth, and perhaps also of fatness, but I

do not necessarily think of the proposition that Seth is fat. To think of the proposition p is more sophisticated than thinking that p. In the former episode one refers to the content of the latter sort of episode, but one is capable of referring only to that which one can conceive of, and one is capable of conceiving the notion of the proposition (i.e. possible thought-content) p only if one is capable of forming the thought that p. The notion of thinking that p must then be logically prior to that of thinking of p, not the other way around.

One must distinguish between two levels of thought, unreflective and reflective thought. At the rudimentary level of unreflective thought, the thoughts of a creature are caused by its perceptions and by its other thoughts, the latter simply according to the mechanism of association. The creature here lacks the power to think of - refer to - the contents of its own thoughts and, consequently, to predicate things of these contents. At the level of reflective thought, the creature possesses this power to refer to propositions and also possesses the ability to apply predicates like those of truth, consistency, entailment, probability etc., predicates that apply to propositions. It then becomes possible for the creature to assess a thought, caused by a perception or by another thought (in an associative fashion), in the light of its other thoughts: to check whether it is compatible with them, whether they entail it or make it probable. The outcome of this process is a reasoned or reflective thought on the subject matter. Such a thought admits of degrees of endorsement or assurance - i.e. it is certain, probable, possible - something which is out of the question at the unreflective level.

The complex topic of reflective theoretical thought must be dropped here - otherwise this would quickly turn into a treatise on theoretical reasoning or epistemology rather than on practical reasoning. But I hope enough has been said to make this thesis plausible:

(T2) "a thinks that p" does not entail "a thinks of the proposition p".

If one holds the negation of (T2) together with (T1), one easily slips into a piece of reasoning that goes like this: "a's thinking of the proposition p exhausts the element of thinking to be found in his thinking that p. But obviously, the former does not entail the latter, because the latter involves endorsement of p. Hence, this component of endorsement must consist of something of a nature quite different from that of thinking." Historically, there have been two types of theory of this additional feature: (a) that the differentials are effects on outer behaviour, (b) that the differentials are mental phenomena, different from thinking. Since hypothesis (a) will be scrutinized in 3.3, I shall here restrict myself to a brief comment on (b).

Under (b) we find the traditional theories of an inner episode of assent. Assent is "the taking up of an attitude towards an entertained proposition" ([161], p. 207). Price distinguishes (pp. 204-8) between two components in this attitude: (i) a Cartesian (vide [74], 4th Meditation) element of decision, choice or preference, (ii) a Humean² element of feeling, in Price's case a feeling of confidence. However, as Price himself immediately goes on to explain (pp. 208ff), there are cases of thinking-that where neither (i) nor (ii) is satisfied: where we accept something in an "unquestioning manner", i.e. Cook Wilson's "being under the impression that ...". (I think this distinction corresponds fairly closely to mine between the two levels of thought.) But even if we restrict ourselves to clear cases of reasoned thinking-that on the basis of evidence, an analysis in terms of (i) and (ii) will not do.

Both suggestions to the effect of (i) and (ii) share a crucial defect: that of circularity. It will be a theme of this work that practical notions such as wanting, preferring (i.e. wanting something more than something else), intending and deciding presuppose the notion of thinking-that. Furthermore, those of these practical concepts which are genuine mental episodes are thoughts of a certain content, e.g. (as I shall maintain in "Conclusions and Limitations" under (1)) a decision is a thought the content of which finds expression in the form "I shall ...". Thus, if thinking that p is thinking of p plus a decision-operator shall - so that thinking that p is represented by shall p (whatever that may mean) - then this decision (being itself a thought-that) ought to be analyzable as a compound of entertaining shall p plus the addition of a second-order decision-operator. But this obviously starts an infinite regress. As concerns emotions or feelings, I will comment briefly in 3.3.1 on the circularity of bringing them in. Let me here merely note the implausibility of the supposition that one particular feeling - the feeling of confidence - is present in varying degrees whenever we have a reasoned thought. Not only are there cases when it seems far-fetched to claim that we feel anything at all (cf. [157], p. 132), but there are also cases where other emotions are felt with such intensity that there seems no room for any feeling of confidence (whatever that precisely is): suppose someone firmly establishes the conclusion that he has swallowed a deadly poison and will die within a minute, then the intense terror or despair he might feel will in all likelihood quench all other feelings - especially a feeling of conviction, which Price characterizes (pp. 288-9) as a feeling of calmness or tranquillity! Hypothesis (b) then turns out to be unacceptable when spelled out, but this is not to be wondered at, because the question it sets out to answer is a spurious one, arising from a misguided rejection of (T2).

To think that p must not be confused with saying to oneself in "inner speech" that p. The latter, just as saying that p aloud (to oneself or others), is compatible with not believing that p. The so-called pep-talk of, e.g. a sportsman, often takes the form of a silent soliloquy, aimed at persuading himself of something, e.g. that his chances to succeed are greater than he is inclined to believe. Saying that p to somebody, whether aloud or silently, is, in contrast to thinking that p, an action, something that often is done voluntarily (cf. [180], pp. 151-2; [181], pp. 106-7). In the case of overt talking one brings it about that certain sounds are formed by one's organs of speech (sounds that are calculated to be heard and understood by a certain audience); in covert talking one presumably brings about - for the same purposes as one produces real sounds when talking loudly to oneself - that one has "images" (for this term, see 3.4) of certain speech-sounds as uttered by oneself. These acts of saying something - of actual and imagined saying - clearly do not involve belief in what is said. (Nor must the thought that p be confused with the result of the internal act of saying something to oneself, with the image representing oneself as saying "p" or that p.)

Zeno Vendler makes - quite consonantly with the Cartesian inclination to assimilate thinking-that to an action - the mistake of supposing that if thinking-that is carried out in some code or other, thinking-that must be more or less the same as talking silently to oneself: "if thinking needed a code, consisting of words or other symbols; if, in other words, thinking were something like talking to oneself silently ..." ([207], p. 44; my italics). It is not to be disputed that Vendler is able to derive bizarre corollaries from the former supposition if it is equated with the latter; but this only demonstrates the absurdity of the equation. For the alternative to the view that thinking requires a code - the thesis Descartes expressed thus: "in reasoning we unite not names but the things signified by the names" (quoted approvingly by Vendler, p. 171) - is certainly false. It would license unrestricted substitution of co-designative terms for particulars and universals in the context "thinks that ...", and this is absurd: from the fact that I think that the servant is wearing a deep-blue necktie, it does not follow that I think that the secret lover of the countess is wearing a necktie of his mistress's favourite colour, even if the servant = the countess's lover, and deep-blue = the countess's favourite colour.³ The only way to account for this phenomenon (further discussed in 3.2) seems to be to assume that thinking is carried out in some code.⁴

Thinking that p must also be kept apart from thinking about (thinking it over) whether p or -p. The latter may be an action,⁵ i.e. the action of causing

oneself to begin to have - or to keep on having - thoughts-that with a content pertinent to the question of whether p or -p, although sometimes one begins to think about a certain topic as e.g. a result of something one perceives. Perhaps the analogous relation between looking-at and seeing-that can be helpful in clarifying these relations. Sometimes one's gaze is simply drawn towards an object one sees peripherally; the object catches the eye, as the saying goes. At other times the focussing upon an object is something one brings about. But even in the latter case, if one sees that the object has such and such properties as a result of this focussing, this seeing-that is not an action, but the effect of an action. Similarly, thinking that p may (but need not) be the effect of (an action of) turning one's thought or attention to the topic whether p or -p, but it is not itself an action.

Thus, this thesis contributes significantly to the characterization of thinking-that:

(T3) To think that p is not to act.

To my knowledge, Hugh McCann is the only analytical philosopher who has recently expressed dissent from (T3). For him thoughts are acts, though "(u)nlike acts of moving a finger or flexing a muscle, thoughts do not have results" ([125], p. 463). McCann must then reject the view of action I have espoused, that an action is a causing. Since he does not offer anything in its place, his thesis that thoughts are act(ion)s is left obscure. In support of his contention, he argues that "people are held responsible for thoughts" (p. 472). But even if this is correct, it is insufficient to establish McCann's conclusion, for it is far from clear that we are held responsible only for our actions: e.g. a soldier can be held responsible for falling asleep on guard (or so it seems), yet falling asleep is no action. Perhaps, it may be countered, the soldier is held responsible because there is an action in the offing - preventing himself from falling asleep - which he omits. But this suggestion can be made with equal right in the case of thoughts.

2. Let us now briefly consider the converse of (T1), to wit

(T4) "a believes that p" entails "a thinks that p".

It has been repeatedly asserted, especially since the publication of Ryle's The Concept of Mind, that a belief is a disposition. I agree with this; more precisely, I accept Wilfrid Sellars' view (expressed with some reservations in e.g. [180], p. 154 and [183], pp. 514ff) that a belief that p is a disposition to think that p. From this it follows that (T4) is false, if it affirms that a cannot believe that p at t without thinking that p at t. But I am prepared to defend (cf. [108], p. 191) a slightly weakened version of (T4), namely

(T4*) "a believes that p at t" entails "a thinks that p (or something explicitly entailing that p) at t or has thought p (or something etc.) at a time earlier than t".

The parenthetical qualifications are necessary in order to take care of the fact that one can be said to believe certain propositions that have never entered one's mind, e.g. a man can be said to believe that birds do not wear boots, although this is something he has never thought. The reason is that it is an evident corollary of what he has thought.⁶

Very different characteristics have been called dispositions. For instance, Ryle classifies ([172], p. 43) both to be brittle and to be a smoker as ("single-track") dispositions. Yet it is evident that they are very different properties: a thing can be brittle without ever breaking, but a person cannot be a smoker without actually smoking - the fact that someone would smoke if ... is surely not enough to make him a smoker! To be a smoker is rather to smoke regularly. Now, is having a belief to be likened to being brittle or to being a smoker? I think it falls somewhere between them. Just as one cannot be a smoker if one has never smoked, so one cannot believe that p if one has never thought something explicitly entailing that p. On the other hand, one remains a believer of that p even if one does not regularly think that p. Consequently, I would suggest a definition of belief that runs like this:

(T5) a believes that p iff: a thinks (something explicitly entailing) that p or is in a state which (a) has been created by a thought (explicitly entailing) that p and which (b) manifests itself in a thought that p if a thinks about a topic to which whether p or -p is relevant.

Two reservations must be attached to this definition. First, by "belief" is sometimes meant an attitude that on a scale of confidence falls somewhere between a conviction, on the one hand, and suspicion or surmise, on the other. To obtain a definition of this notion we must substitute "a thought that it is probable that p" for "a thought that p" in (T5). This notion (which I myself will frequently employ) applies only to the level of reflective thought. Second, there are niceties, like temporarily forgetting something, the elucidation of which would necessitate further elaboration of the triggering circumstance mentioned in (b).

If (T5) is to be an explication of the view that a belief is a disposition, the "realist" conception of dispositions D.M. Armstrong has argued for ([5], chap. 6, sec. VI; [6], chap. 2, sec. II) must be correct. According to Armstrong, to say that a object o is brittle is to say that o is in an internal state (which in fact, intrinsically described, is a certain molecular-bonding) such that o will break if struck (and not surrounded by protective

material etc.), not only (as the "phenomenalist" account goes) that o will break if struck (etc.).

Armstrong's view seems very plausible to me, although his argument in favour of it does not appear to establish its conclusion. A line of reasoning which supports Armstrong's analysis is this. The statement "o is brittle now" cannot be logically equivalent to "If o is knocked now (even comparatively lightly) and the circumstances are not such as to neutralize the impact of the knock, o will break". For the latter can be true when the former is false: suppose e.g. that o is delicately poised high above the ground and that a light push will make o fall. Clearly, the construction of such counter-instances cannot be obstructed by adding further specific negative restrictions on the external circumstances (e.g. that o must not be so placed that it could fall), for there are an indefinite number of such circumstances (e.g. a light knock puts o within the range of a falling sledge-hammer). It seems rather that such counter-examples must be excluded by demanding, positively, that o possess some feature F such that the fact that o instantiates F, together with the external facts that o is knocked, even lightly, and that nothing neutralizes the impact of the knock, constitutes a full sufficient causal condition for o's breaking. Now if it is granted either that only internal features of o could play the indicated causal role or that this is a pervasive belief among speakers - and at least the latter seems to be a modest assumption - it looks as if Armstrong's view of dispositions is correct. This view also appears to be supported by certain facts about the way we handle disposition-descriptions, e.g. that these can be used when effects are cited: "The freezing of the material caused it to become brittle" (cf. [11], p. 112). On these grounds, I will in what follows assume the realist view to be true.

3.2 The Content of Propositional Thinking.

1. It is not uncommon to find philosophers treating that-clauses of propositional attitudes as akin to quotation, e.g. "a says that p" or "a thinks that p" as to be illuminated with the help of "a says 'p'" and "a thinks 'p'", respectively. Wilfrid Sellars is an example of such a philosopher (e.g. [182], chap. III; cf. also [64]). The success of this analysis depends on the soundness of two moves.

(i) On the broadening of conventions of quotation so that the truth of e.g. "a says 'It is raining'" does not require that a utters the particular English sentence "It is raining"; a sentence in another language must do just as well as long as it is synonymous. (Cf. Sellars' device of "dot-quotes".) For the truth of "a says that it is raining" does not entail that a speaks any

particular language.

(ii) That "a says 'p'" can be true only if a (on the particular occasion) uses the sounds corresponding to "p" with meaning, i.e. according to the rules of the language the speaker speaks. This, according to Sellars ([184], p. 187), distinguishes "saying '...'" from "uttering the sounds '...'".

Obviously, that a says that p entails that he uttered some sounds meaningfully.

For my own part, I am not willing to grant (ii). It seems to me that a parrot can be described as saying "Open my cage", even though these sounds have no meaning in its beak. What is necessary for the appropriateness of the classification of something as saying "...", - as opposed to producing the sounds "...", - is only that what fills the blank can be meaningfully used (i.e. is words, sounds with meaning), not that they are meaningfully used on the particular occasion. Hence, "a says 'p' (or something synonymous)" does not entail "a says that p" - for the latter does indeed imply that we have a piece of meaningful verbal behaviour.⁷

I will take "a says that p" as being logically equivalent to something like "a utters some sounds in accordance with some convention (public or private) to mean that p". The corresponding thesis for "a thinks that p" will not run completely parallel, since thinking-that is (a) no action and (b) an inner occurrence, which makes it idle to insist on the propriety of the code. I will understand it as being equivalent to "Something occurs in a's mind with the meaning that p".

To this proposal it will probably be objected that it is unenlightening, because the force of constructions like "meaning that p" is not one whit clearer than the definienda. The objection is reasonable; so let me try to clarify the meaning of meaning somewhat. Take a simple declarative sentence of the form "b is F" - how shall we interpret the statement that it occurs with the meaning that b is F? It will carry this meaning (on a particular occasion) iff "b" refers to b, "F" signifies the universal F, and "is" serves its function of joining these terms together in the predicative fashion (all on the particular occasion). But then a's thinking that b is F involves some sort of relation between a and b; "a thinks that b is F" entails (but, as we shall see, is not entailed by) "a thinks about b that it is F". According to this interpretation, definite descriptions appearing in contexts like "thinks (says etc.) that ..." satisfy one test of Quine's for being referential (see [164], §§ 30-2, [165] and [166]), namely the test of existential generalization, for "a thinks (says) that b is F" entails "There is an x such that a thinks that x is F".

It might be objected that only thought or belief de re is a relation

between the thinker and another existent entity, but that there is also thought or belief de dicto - which is not about anything in the world - to reckon with. Now, if the drawing of this distinction is meant to affirm that a sentence like

(1) a thinks that John's present wife is blonde

is ambiguous due to the fact that "thinks that" is ambiguous,⁸ sometimes (so to speak) eclipsing the reference of "John's present wife", sometimes letting it through, then I must protest. I see no reason to assume such an equivocality of "thinks that".

However, some of the intuitions behind the distinction can be salvaged, for (1) is indeed ambiguous - only the ambiguity has to be relocated to the content-proposition. In other words, the sentence

(2) John's present wife is blonde

can be interpreted in more than one way. Imagine (2) being put forward exclusively on this basis: John changes wives frequently, but he always has exactly one wife, and he is only attracted by blondes. The speaker has no personal knowledge of the woman presently in question. (2) is then replaceable by:

(2*) John's present wife, whoever she is, is blonde (or, more perspicuously, "Whoever is presently married to John is blonde").

But if the speaker of (2) is acquainted with John's present wife, "knows who she is", he cannot very well use (2*) instead of (2), since that would be to deny his personal knowledge of the woman. Only in this second situation - where addition of a "whoever"-formula is impossible - is the statement about an individual, known to the speaker, that is John's present wife. In the former case, it is tempting to equate the statement with a quantificational one to the effect that there is exactly one individual such that it is presently married to John and ..., but this does not appear to be entirely correct (this is displayed when the sentences are negated). Even (2*) is about the individual who is John's present wife, though the woman is here (virtually) unknown to the speaker.

This ambiguity of (2) is preserved when "a thinks that ..." is tacked onto it. But if what I just said is true, the thought remains de re, i.e. about somebody (known or unknown), whichever of the two interpretations of (2) is chosen.⁹

This account prompts the question: which are the criteria for personal knowledge, for "knowing who someone is", i.e. when is one in a position to use a definite description as a phrase that refers to someone known rather than merely as a "whoever"-formula? I do not claim to know the answer to

this, but I believe that what David Kaplan [110] has to say concerning the notion of a name's "representing" something to someone is illuminating (though in need of further refinement). In particular, I believe he is right in stressing the genetic side: a thing must have played a causal role (which may be very indirect) in a person's acquisition of some beliefs, if he is to have beliefs about that thing as something known. Fortunately, the problem need not be solved here. All I need to do is to stipulate that whenever I use definite descriptions in rendering the contents of thoughts, they are intended to be taken as referring to known individuals, i.e. as not being susceptible to qualification by a "whoever"-clause.

Quine usually runs together two tests for referential positions, viz. the tests of existential generalization and of substitutivity salva veritate of terms that purport to refer to the same individuals. To my mind, however, these criteria are not co-extensive. When Quine presents the latter criterion (in [164], p. 142), he uses it as a test for purely referential positions, i.e. for positions where a term's sole function is to specify its referent. But suppose a term in a context does double duty: it refers to an item all right, but also indicates a favoured way of referring to it; suppose e.g. that "b" in the sentence "a thinks that b is F" serves the functions (i) of referring to b, and (ii) of indicating how a refers to b in his thought (i.e. by using "b"). Then of course unrestricted substitution of co-referring terms cannot be allowed, since there are ways of referring to b a is unaware of. But what about the test of existential generalization? It seems that this operation is still legitimate, for if "a thinks that b ..." is equivalent to "a thinks about b, referred to as 'b', that it ...", then existential generalization will yield "a thinks about something, referred to in some way, that it ..." - and this inference is sound. So, the outcomes of the two tests are not identical: existential generalization lets in not only purely referential terms, but terms that at least refer.¹⁰

Now I propose this definition:

(T6) a thinks that b is F =

a thinks about b, identified by (what translated becomes) "b", that it has the property F, signified by (what translated becomes) "F", i.e. something occurs in the mind of a that refers to b, by (what translated becomes) "b", that signifies F, by (what translated becomes) "F", and that joins these terms in the predicative fashion.

"Translation" is here intended to include - beside the procedure of reformulating something expressed in one natural language in another - the shifting of index-words, especially personal pronouns, made necessary by the change

of personal point of view, and substitution of intra-language synonyms, perhaps of differing stylistic value.

The context "thinks that ...", so interpreted, admits of existential generalization, but unlimited substitution of co-designative terms is illegitimate: substituends picked from outside the range of the thinker's thought yield falsehoods.

However, Saul Kripke has shown [115] that a formula like (T6) is not entirely correct. He has argued (pp. 254-63) that it is possible that a person, who speaks both English and French, uses "London" and "Londres" to refer to what is in fact the very same city, but without realizing this. Such a person might sincerely assent to both "London is not pretty" and "Londres est jolie". If so, (T6) justifies the attribution both of the thought that London is not pretty and the thought that London is pretty to the person. This, however, seems unacceptable since the thoughts contradict each other blatantly, which the assertions do not.

The counter-intuitive consequence can be blocked if this stipulation is added to (T6): if a uses both the proper name "n" and the proper name "n*" to refer to the same entity, "n" is adequate as a translation of "n*" in the context of a's propositional attitudes only if a realizes that "n" and "n*" refer to the same entity. But it must be shown that this stipulation is not merely ad hoc. Consider two descriptions - i.e. two terms that clearly have connotation as well as denotation - d and its correlate in another natural language, d*. If a uses d and d* to refer to the same entity, but without being aware of the sameness of reference, it follows that he does not treat d and d* as synonyms - and hence that, in his idiolect, d is no translation of d*. But due to the semantic poverty of proper names - their scarcely having connotation in any straight-forward sense - this reasoning does not apply to them. In their case, the requirement of "translatability" does not entail that the person involved realizes the sameness of reference; this must be separately added. If this constraint is violated, a proper name, e.g. "Londres", cannot be translated into its equivalent in another language.¹¹

Even with this qualification, (T6) may not be exceptionless: in some contexts we apparently allow substitutions even outside the cluster of descriptions known to the thinker. Someone could perhaps say e.g. "A famous art-dealer believed this fake to be worth at least 5.000.000 dollars". But I believe these exceptions can be accounted for if one takes one's departure from (T6): it might be suggested that in these contexts a qualification of the form "but of course the thinker does not identify the relevant items in this way" is implicitly understood. Thus, I am inclined to regard (T6) as

defining the paradigmatic or central sense, but I need not insist on this; it suffices if it is acknowledged to be a possible sense.

2. Thinking that $\underline{p}$, then, is possible only if there is such a thing as having the meaning that $\underline{p}$. This will help me to make the notion of a proposition - i.e. a possible (true-or-false) content of thinking-that - a bit clearer.

Some contradictions cannot fill out "thinks that ..." for the reason that they have no meaning. For instance, there is no meaning that $\underline{p} \ \& \ -\underline{p}$, for the meanings of "&" and "-" are such as to exclude this construction. If someone in all sincerity uttered sentences of the form " $\underline{p} \ \& \ -\underline{p}$ " and did not accept some non-contradictory formulation (e.g. "The case at hand lies on the borderline between $\underline{p}$ and $-\underline{p}$ ") as adequate for what he intended to convey, then this would prove that he did not master the use of both "&" and "-".

However, not all contradictions are of this explicit sort, i.e. of the sort where one perceives the contradictory character the moment one grasps the meanings of its components. A person cannot be said to have understood the meaning of "triangle" if he has not realized that it designates a closed three-sided figure, but (let us at least assume this for the sake of illustration; it is hard to determine what entailments are part of a term's meaning) he need not grasp that it designates a figure whose area is $\frac{b \cdot h}{2}$ or whose angles have the sum 180° .¹² If so, there is the meaning (proposition) that the area of this triangle is not $\frac{b \cdot h}{2}$, but not the meaning (proposition) that this triangle is not a three-sided figure; and it is possible to think that the former, but not the latter, is the case. However, once one begins to think that if something is a triangle then its area is $\frac{b \cdot h}{2}$, one can no longer think that this is a figure the area of which is not $\frac{b \cdot h}{2}$, while one thinks that it is a triangle - for such is the meaning of "if-then". (People who have realized the analyticity of this if-then statement may use "triangle" with a richer meaning, so that for them "The area of this triangle is not $\frac{b \cdot h}{2}$ " has no meaning.)

So, some contradictions can, others cannot, logically, be put into the blank of "thinks that ...": an explicit contradiction cannot, because its contradictory character is apparent once the meanings of its constituents are grasped, and this grasping is a necessary precondition for thinking-that. This implies a certain view of the sense in which the law of contradiction is a "law of thought". It is not a normative law regulating the process of thought, if by a "normative law" one means a law one can consciously or intentionally violate or, alternatively, a law obedience to which is a matter of choice.¹³ Rather, the law expresses a conceptual truth about the notion of thought; it lays down a restriction on what is to count as thinking that something is the case. It is logically necessary that if $\underline{a}$ thinks that if $\underline{p}$ then $\underline{q}$ and that $\underline{p}$,

he must tend to think that q, i.e. he must think this if the question whether q or -q is something that catches his attention. If he is not disposed to think that q, he does not understand the customary meaning of "if-then" constructions, which is necessary for an oratio obliqua ascription of a thought of conditional content.

For the normative view the question why people do not (at least as a rule) believe what they regard as contradictory becomes embarrassing. People can have no reasons for this, since every piece of reasoning clearly presupposes the law of contradiction. But is it then a brute, unexplainable fact or a fact to be explained by some non-rational fact, say a feature of the human (mammal) brain? Neither alternative seems acceptable. I mention these contrasting theses about a rule of theoretical reasoning chiefly because I will later (in 6.1) partake in a similar dispute about the rules of practical reasoning. There I will, as here, argue against the normative view and for the conceptual standpoint.

An explicit contradiction is not a proposition, since it cannot form the content of thinking-that. But, I think, we might just as well exclude explicit logical truths from the class of propositions, that is, require that "p" expresses a proposition only if both "a thinks that p" and "a thinks that -p" are logically possible. For an explicit logical truth can hardly constitute a reason for action: if "If I bring about p, then q" expresses something whose analyticity is obvious, and a brings about p only because he thinks this, then he has clearly brought about p for its own sake, not for a reason.

3.3 Thought and Behaviour.

1. In this section I shall try to establish a negative point concerning the relation between propositional thinking and (observable) behaviour, namely that the former is not to be defined even partly in terms of the latter. I will put the thesis to be defended in this way:

(T7) The phrase "to think that p" is a rigid designator of a certain (type of) mental property.

In the vocabulary of Saul Kripke ([114], esp. pp. 269-70) a designator d is a rigid designator of an item i iff d (so long as it is used as it is in our language) designates i in every possible world that includes i.

A champion of functionalism or the causal theory of mind will dispute (T7).¹⁴ The philosopher who has carried out the programme of the causal theory in greatest detail is D.M. Armstrong in [5]. Armstrong states his general view thus:

"The concept of a mental state is primarily the concept of a state

of the person apt for bringing about a certain sort of behaviour ... In the case of some mental states only they are also states of the person apt for being brought about by a certain sort of stimulus." (p. 82)

Given my preference for facts (property-instantiations) in the roles of cause and effect, I would restate the view in this way: a mental property is, primarily, a property the instantiations of which are apt to cause a certain sort of behaviour and, secondarily, a property the instantiations of which are apt to be the effect of a certain sort of stimulus.¹⁵ If this is correct, designators like "to think ...", "to perceive ..." etc. are non-rigid designators, for a property the instantiations of which in this world have certain causes and effects will not in every possible world have these same causes and effects. Alternatively phrased, they are descriptions which identify by specifying some contingent relational characteristics. Armstrong believes that in point of fact it is states of the central nervous system that are the effects of stimuli and the causes of behaviour. According to this, a rigid designator of the properties mental terms designate non-rigidly will be fetched from the terminology of neuro-physiology.

There is a rather naive - but still forceful - objection to this conception of a mental state: the conception leaves out something that is quite distinctive to mental phenomena, namely that they have intrinsic qualitative characteristics, that sense-impressions have e.g. colour-properties and thermal qualities, that thoughts have content. Armstrong touches upon the difficulty when he notices ([5], pp. 118-9) that not every state apt for the production of behaviour is a mental state.¹⁶ He tries to meet the objection by requiring that the cause must reach a certain degree of complexity ([5], pp. 119-20, 252-3). But this is clearly inadequate: it is simply not the case that, whenever the physical causes of some bodily change reach a certain degree of complexity, they can be characterized in mental terms. The brain's control over bodily processes like sweating, the beating of the heart, breathing etc., is a complicated matter, but it does not follow that the control is mental.

If this is conceded, the causal theory can be modified in two ways. (a) Reference to other mental states can be brought into the characterization of a certain mental state, so that it is no longer identified exclusively by its relations to (observable) input and output.¹⁷ Towards the end of this section I will consider such an explication of the concept of belief. (b) The purely causal standpoint can be given up and some mental state bestowed with some intrinsic mental property. I will suggest that Armstrong may have to take this line in his theory of belief.

This theory is more fully developed in [6]. I will here concentrate on

Armstrong's analysis of what is for him the fundamental type of belief, namely beliefs concerning particular matters of fact (in contrast to beliefs in unrestricted universally or existentially quantified propositions). Such a belief is, according to Armstrong, an internal state of the believer.

As hinted at in 3.1.2, dispositions are, for Armstrong, a species of state. To ascribe brittleness to a thing is to assert that it is in a certain internal state, which however one uses a non-rigid designator to identify, i.e. the state is picked out by some of its relational characteristics (what would happen if). A rigid designator of the same state would be something like: "molecular bonding of structure M".

However, Armstrong discerns the following differences between a disposition and a belief (of the indicated type):

"(i) Manifestations of dispositions are 'stimulus-dependent'; manifestations of beliefs are not. (ii) Dispositions, if manifested, are manifested in only one sort of way; beliefs, if manifested, are manifested in indefinitely many ways. (iii) While attribution of a disposition to an object attributes a state to the disposed thing, the state need be credited with no particular structure; but belief-states must be credited with a structure which corresponds to the proposition believed." ([6], p. 21)

The import of (i) and (ii) is that it is part of the concept of a paradigmatic disposition like brittleness, as opposed to a belief, that the triggering circumstance and the manifestation are of a certain kind: being struck and breaking, respectively. For my part I am inclined to deny this contrast; the stimulus and manifestation descriptions appear just as determinate for a belief that p, namely, when it becomes relevant for what the believer is thinking whether p or -p and thinking that p, respectively.

In discussing (ii), Armstrong makes the puzzling remark that due to the multifarious nature of the possible manifestations of a belief they "can only be identified as manifestations of the belief by reference back to the belief" ([6], p. 17). This sounds like an argument against Armstrong's own view. The illustration he provides strengthens this impression. He writes that saying "p" can count as a manifestation of a belief that p only on the condition that conventions lay down that a belief that p is to be expressed by the sound "p" (and, I would add, the speaker knows and wants to follow these conventions). This appears to be a charge (which I would sympathize with) that reference to verbal manifestations in the definition of belief generates a circle (cf. [157], p. 128). But what Armstrong has in mind is probably a point that is echoed in (iii), that it is part of the concept of belief (in contrast to the concept of e.g. brittleness) that the state causally responsible for the manifestations has a certain internal structure, a structure

which somehow corresponds to the proposition believed. The quoted remark is presumably meant to assert that a piece of behaviour can only be identified as a manifestation of a belief that p if it is found to result from a state with a structure corresponding to that p.

This state, apart from its effect on behaviour, cannot be classified as a belief; it must be what Armstrong classifies as a mere thought (i.e. entertainment of the proposition) that p. It seems to be Armstrong's view that there are differences in two respects between a belief that p and a mere thought that p. (a) A mere thought is a mental occurrence ([6], p. 36), a belief is only occasionally occurrent. That a belief is occurrent means that it is currently causally active; Armstrong uses "thought" in a second sense - which comes close to my use of the term - to designate occurrent belief ([5], pp. 343-6). So, we still have to find a difference between mere thoughts (entertainments of propositions) and thoughts proper (i.e. occurrent beliefs). (b) "the belief-state is, and the /mere/ thought is not, a potential cause or inhibitor of action" ([6], p. 72). The difference between mere thoughts and proper thoughts, then, consists in the causal efficacy alluded to in (a). But both beliefs (thoughts proper) and mere thoughts have the same internal structure, a structure which corresponds to the propositional content:

"We believe this proposition /that there is a red thing over there/ if, and only if, we are prepared to act towards the thing, or supposed thing, in a 'red-selective' way, if we should desire so to act. The mere thought that there is a red thing over there would be something in our mind which had the same structure ... but which was not, even in conjunction with suitable dominant desires, a cause of such selective behaviour" ([6], p. 71).

Armstrong sums up his view of the relationship between mere thought and belief in this way: "beliefs are thoughts plus. The plus is the causal role which beliefs play in behaviour" ([6], p. 72). So, it is apparently Armstrong's view that it is not part of the concept of a mere thought to be causally effective in behaviour.

Yet a mere thought is undoubtedly a mental state, and according to Armstrong's general conception of a mental state that is equivalent to being (primarily) a state apt for bringing about behaviour. It will not do to reply that mere thoughts at least have a causal influence on introspection, for Armstrong distinguishes ([5], e.g. p. 345) between mental occurrences and conscious mental occurrences, and only the latter are objects of introspection. (Nor could the causal influence observed in introspection be the influence on introspection itself.) Armstrong appears inclined ([5], pp. 348-9; [6], pp. 71-3) to appeal to a resemblance between the internal states which reflects the proposition in the case of both mere thought and thought proper. But

unless these states are conceived as being mental in some intrinsic or non-causal sense, this will not show that mere thoughts are similar to thoughts proper in that both are mental properties. That is why I said that Armstrong may be forced to abandon the pure causal standpoint.

Assuming that he does leave it, his theory will be substantially identical to that of R.B. Braithwaite in [31].¹⁸ Braithwaite here distinguishes (p. 30) two components in a belief that p: (i) a mental episode of entertainment - or a disposition to entertain - the proposition that p, and (ii) a disposition "to act as if p were true". In any event, the criticism to follow will not turn on how the mentality of entertainment is to be understood. For reasons of space, I will only present the considerations that seem to me to weigh most heavily against the theory.

(1) According to these accounts "to think (properly) that p" is roughly synonymous to "to entertain the proposition that p when an instance of this is apt to cause behaviour". (Hence, it will be a non-rigid designator, for the instances of entertainment that tend to cause behaviour are not the same in all possible worlds.) Now, according to the Armstrong-Braithwaite view, it is misleading (though not downright false) to say that thoughts proper cause behaviour, for this may create the misapprehension that (in this view) entertainments of propositions do not have the same causal power, that "to think (properly) that p" designates a property the instantiation of which is a causal condition of behaviour, a property that "to entertain p" does not designate. But on the position under consideration the assertion that entertainment of p causes behaviour is just as true as (and less misleading than) the assertion that thoughts (proper) that p cause behaviour, since to characterize something as a thought (proper) that p is merely to say that it is an instance of an entertainment of p which has the relational property of tending to cause behaviour; and this relational property is not a further property that is causally relevant for the production of behaviour.

It is obvious, however, that we ordinarily treat "to think (occurently believe) that p" as designating a causally relevant property that "to entertain p" does not designate. Suppose that a entertains p. Clearly, this would seldom, if ever, cause a to act in a way that presupposes that p is true, for him to achieve what he wants. As a rule, one dares not act on a proposition until one has acquired belief in it. a acts in a way that presupposes the truth of p because he begins to believe that p; it is not the case - as it would be on the Armstrong-Braithwaite view - that a's entertainment of p becomes a belief by his acting on it. It is possible to explain the occurrence of an action at t by pointing out that the agent at t - perhaps by acquiring

evidence - changed his attitude towards p from entertainment to belief. But the criticized doctrine does not permit this, since according to it the characterization of something as becoming a thought (proper) is only a restatement of the explanandum, that the entertainment has acquired the capacity to produce behaviour; it does not explain why this suddenly has become so. Consequently, in situations like these, it leaves the genesis of the action wrapped in mystery.

(2) Ryle advances the thesis that "believe" is a "dispositional word" ([172], p. 118). He immediately adds that it stands for a "highly generic or determinable" disposition (pp. 118-9; cf. 44-5), i.e. that it stands for a disposition that manifests itself in multiple ways. An example of his illustrates this:

"to believe that the ice is dangerously thin is to be unhesitant in telling oneself and others that it is thin, in acquiescing in other people's assertions to that effect, in objecting to statements to the contrary, in drawing consequences from the original proposition, and so forth. But it is also to be prone to skate warily, to shudder, to dwell in imagination on possible disasters and to warn other skaters" ([172], pp. 134-5).

Actually, Ryle could go on expanding this list of manifestations forever, for the truth is, as critics have pointed out (e.g. [86], p. 8), there is no set of behavioural patterns characteristic of a given belief. How a person behaves depends not only on his beliefs, but also on his wants, aims etc. (A third relevant variable is his ability, but let us ignore this for the moment.) If, for instance, the skater wants himself and others to drown, he will obviously neither skate warily, nor warn others.¹⁹

So, Armstrong is right in insisting (e.g. [6], p. 71) that a belief is only apt for bringing about a certain sort of behaviour when taken in conjunction with certain wants, purposes etc. But it then falls to his lot to provide a definition of these conative concepts without appealing to the concept of belief - otherwise a circle will ensue. This task of analyzing the conative concepts is undertaken in [5], chap. 7. In fidelity to his causal theory, Armstrong there claims that purposes, intentions etc. "are states with causal powers: powers to initiate and sustain trains of physical or mental activity" ([5], p. 136). Furthermore, they are "information-sensitive" mental causes ([5], p. 139), i.e. mental causes whose "operation is controlled by what the agent believes to be the case" ([5], p. 150). For example, behaviour caused by a purpose is, when developing, modifiable in the light of information feeding back via the channels of perception; there are also beliefs that extinguish behaviour, namely beliefs that the objective has materialized or that it cannot materialize. According to this, to find out the purposes of an agent is to find out which of his beliefs direct his behaviour.

It is now evident that Armstrong does indeed move in a circle. Armstrong himself is aware of this circularity in his account, but tries to defend it by arguing ([5], pp. 251-4; [6], pp. 66-7) that the circularity is not vicious:

"this circularity is not vicious. It simply points to the fact that purpose and belief are 'package-deal' concepts, concepts, that is to say, which must be introduced together or not at all. (A simple example of such a concept-pair is the concept of husband and wife.)" ([6], p. 67)

However, this "simple example" - and others Armstrong gives - reveals that he has not correctly diagnosed the problem: he has confused interdependence of connotations of terms and interdependence of their applications. A world in which "husband" has application but in which "wife" has never been applicable (or vice versa) is not possible. Still, "husband" can be defined without reference to "wife", e.g. as "man married to woman". Hence, we have here only the second type of dependence. Now, Armstrong appears to take the interdependence his theory assumes between "belief" and "will" as being only of this second sort: this is revealed not only by the parallel drawn in the quotation just given, but also by his arguing that when a system of causes reaches a certain degree of complexity the concepts of belief and of will become applicable together, as designating "two sides of one hugely complex process" ([5], p. 252). But if this interpretation is correct, Armstrong here simply fails to grasp the nature of his theories of belief and will, for these clearly make the connotations of "belief" and "will" interdependent. To point out the innocuousness of interdependence of application does nothing to meet the charge of circularity, for it concerns the interdependence of connotations.

Circularity need not always be an objectionable feature. Ultimately, it is perhaps inevitable. But the larger the number of definitional links necessary to create a circle, the greater the information concerning conceptual interrelations. In Armstrong's case, however, the circle is very tight: two terms figure directly in each other's definitions. Such a theory is, due to its low information-value, acceptable only as a last resort, if no other theory making the concepts not directly interrelated can be sustained.

The concept of thinking is one of the fundamental mental concepts. Hence, if there is a connection between it and another mental concept, the risk is great that the latter concept must be analyzed in terms of thinking. Wanting is, as I shall show later, such a concept. Emotion is, though I cannot argue it here, another. Therefore, one also involves oneself in a circle if one conceives belief, as Price ([161], p. 294), as a multiform disposition that manifests itself - apart from in behaviour - "in emotional states such as hope and fear; feelings of doubt, surprise and confidence". (Cf. also [172],

pp. 134-5). For what distinguishes an emotion from a sensation? Certainly, one point of difference is that it is part of the concept of an emotion to be related to (some would say: caused by) a thought. For instance, how define surprise - "perhaps the one /feeling-state/ which is most closely connected with belief" ([161], p. 276) - except as the reaction one shows when one learns that something one firmly believed is false?

(3) The theory is not restrictive enough: the class of propositions believed will be much greater than we would ordinarily admit. For instance, on all accounts there should be something absurd or remarkable about a belief in an explicit contradiction - but on the theory under examination this is not so. There is nothing exceptional about simultaneously entertaining both p and -p (e.g. asking oneself whether p or -p). And suppose we now accept the doctrine that entertainments of propositions, together with desires, cause behaviour. Then if there is a possible desire, which conjugated with the entertainment of p, produces a certain pattern of behaviour, there is no reason why there should not be another possible desire, which in conjunction with an entertainment of -p, brings forth another pattern of behaviour. But if so, it follows from the view scrutinized that there is simultaneously in mind one thought (proper) that p and one that -p. Indeed, on this view it is hard to understand how there can be cases of mere thought: if there are possible circumstances in which entertainments of some propositions cause behaviour, why should there not be such possible circumstances for entertainments of all propositions? On the other hand, if one requires that there be an actual desire, actual circumstances in which an entertainment tends to cause behaviour, the theory obviously becomes too narrow, since every moment a person thinks a lot of things that are irrelevant to his present desires.

The absurdities of Armstrong's theory tend to be effaced by its permission of statements seemingly in accordance with ordinary ways of talking about belief, thus "beliefs are, mere thoughts are not, premisses in our practical reasonings" ([6], p. 74). But the interpretation he puts on them is quite deviant, namely that it is not the case that "belief" designates a property causally relevant for the production of behaviour that "mere thought" does not designate. This, I have argued, is the direct negation of the interpretation ordinarily bestowed on them.

2. It appears that some of the difficulties above are due to the attempt to give a (possibly) internal characterization of belief (as entertainment of a proposition). Suppose this is given up, and belief is determined solely by its causal or functional role: by its relations to other psychological concepts

(like those of desire and emotion), and by its relations to observable phenomena like behaviour, verbal and otherwise. Such a view - defended by Robert Audi [7] and Mitchell Ginsberg ([87], pp. 20ff) - sees the concept of belief as a theoretical concept introduced as part of a theory designed to explain behaviour.²⁰ The significance of a theoretical concept is fixed by a set of laws relating the concept to other theoretical concepts and linking it to phenomena on the observable level.

It is doubtful whether this move creates a more plausible theory. First, the considerations under (2) continue to pester this account. There is the circularity consisting in the fact that belief turns up in the explication of the notions of wanting, having an emotion and asserting. But what is perhaps even worse, given the interdependence of believing and wanting in the production of behaviour, a unique relation to observables for neither of these postulated entities has been specified. If no such relation can be provided, it seems that the principle of simplicity forbids us to uphold the distinction between them, to postulate more than one entity. It is of no avail, I think, to introduce a reference to the stimulus-side, since one here faces a comparable problem of providing unique roles for both beliefs and sense-impressions.

Second, the theory under consideration seems hard put to explain the knowledge one has of one's own current beliefs. An infallible method of becoming aware of what one presently believes on some matter is simply to consider the matter, to survey as many reasons for and against as one is capable of calling to mind, and form a belief on the basis of them.²¹ The question "Do I believe that p?" and "Is it reasonable to believe that p?" must receive the same answer ([75], pp. 90-1). This is why there is no use for locutions like "I don't know whether I believe that p", except as a way of saying that one has not yet arrived at a stable belief regarding p, does not know what is reasonable to believe regarding it; in other words, as way of saying what is more aptly expressed "I do not know whether to believe that p" ([157], p. 130). But according to the theoretical construct view, to ascribe a belief to oneself is to put forward a hypothesis about what lies behind certain manifestations (cf. [87], esp. pp. 54-5). This implies there will be only a difference of degree between ascribing beliefs to oneself and to others ([87], p. 63), while there is in fact a difference of kind. For my own part, I want to explain knowledge of one's own present beliefs as an instance of short-term memory of what one thought the moment immediately before, which will be the same as one is presently disposed to think, since there has not been time for any other thoughts on the topic to intervene. To be sure, it is conceivable that short-term

memory is deceptive, but then reflective reasoning will be impossible, and the question "Is it reasonable for me to believe p?" unanswerable.²²

Third, the highly sophisticated concept of a theoretical concept pre-supposes the concept of belief. A theoretical concept is roughly the concept of an (unobservable) entity postulated to explain certain observed phenomena. Now what is explained is strictly what is (correctly) believed about objects of observation. And to explain these beliefs is roughly to induce other (correct) beliefs to which the former stand in a certain relation (presumably of deducibility). So the notion of a theoretical concept cannot be used to elucidate that of belief without circularity.²³

The conclusion to be drawn is that belief is not a theoretical concept. It might be retorted that my own approach contains the thesis that belief is a theoretical concept, since it refers to a state which underlies the manifestations in thought. However, this is not so, because thinking that p entails believing that p. It is a contradiction to suppose that the existence of a theoretical entity is logically guaranteed by phenomena on the observable - or introspectable - level.

3.4 The Medium of Thought.

1. My objective here is to look into the question whether there is a natural and unlearned medium of thought or whether all thinking is carried out in some conventional and acquired symbol-systems like those of ordinary (commonly, but not here, called "natural") languages. More precisely, I will argue for two theses:

- (T8) There are instances of propositional thinking the contents of which are formulated at least partly in some natural medium, namely that of images;
- (T9) There are instances of propositional thinking the contents of which are formulated altogether in some natural medium.

As will become apparent in chap. 6, (T8) is of importance for the credibility of my theory of acting at will. The plausibility of my theory is enhanced if it is true that images can function as a medium of thought. (T9), on the other hand, does not have any direct bearing on my theory. But it is desirable to touch upon the question of the range of applicability of the terminology of reasons, and for this (T9) is relevant. Not that (T9) in itself is sufficient to establish that reasons can be ascribed to creatures not possessing a conventional language; for this it is necessary that the natural code is rich enough, not only to express some propositional forms, but a rather advanced one, viz. the conditional form. It is dubious whether the natural medium suggested here - that of images - is powerful enough to express any pro-

positions at all, let alone conditional ones. So I am unable to give a clear-cut answer to the question of the applicability of reasons, for though I will indicate an argument for the existence of a natural code of appropriate richness, I am incapable of giving a concrete description of this code.

2. There is a theory of thinking - perhaps associated especially with the names of Wilfrid Sellars and Bruce Aune²⁴ - that tries to elucidate thinking by construing it as an analogical extension of overt speech. The aim of the theory is to show

"that the notion of silent thinking can be seen as an essentially public one, derived from other notions that apply unproblematically to publicly accessible phenomena" ([11], p. 192),

viz. to overt utterances; the relevant notions that are supposed to apply to these are semantical ones like meaning, reference and so on. The derivation takes the form that silent thinking is construed as a theoretical entity on the model of overt speech, as an entity introduced to explain behaviour unaccounted for by the agent's actual utterances. To carry out this programme linguistic behaviour that satisfies these two conditions must be found: (i) it must be sufficiently rich to permit the derivation of the concept of thinking (which means that it must be characterizable in semantical terms), (ii) it should not presuppose the concept of thinking (cf. [182], pp. 71, 155). An ordinary speech-act like assertion does not satisfy (ii): it involves a reference to thought already in that it is an example of an intentional or purposeful action ([182], pp. 73-5), but also in other ways ([11], pp. 191-2). So, the "basic analogy must rather be drawn between silent thought and a proto-version of overt speech" ([11], p. 192). Sellars' name for this rudimentary verbal response, not an action in any qualified sense, is "thinking-out-loud", while Aune's is "proto-assertion".

Now, what is it to think-out-loud or proto-assert? Aune answers:

"From an extensional point of view, what obviously happens in proto-assertion is that an agent emits certain noises while in a frame of mind (state of readiness) to emit certain other noises and perhaps to make certain physical movements, depending on what happens, occurs, or exists in his immediate sensory vicinity" ([11], p. 201).

One further important variable determining which particular noises the speaker is disposed to emit is the specific set of conventions regulating the use of the sounds of the speaker's language; to characterize something as a proto-assertion is "to subsume it under a network of essentially normative concepts" ([11], pp. 204-5). Aune and Sellars picture language as a game, a game governed by three main kinds of normative rules: rules that fix relations world-word (the observing game), rules that fix relations word-word (the inferring

game), and rules that regulate word-world connections (the acting game). The meaning of a word is the position assigned to it in the language-game by these rules (see esp. [178], pp. 327ff). A speaker talks meaningfully to the extent that his utterances (or dispositions to utter) conform to these rules.

If we now compare the rules of different languages, we find certain correspondences between them: e.g. the rules governing "It is raining" in English correspond to the rules governing "Es regnet" in German and "Det regnar" in Swedish. On the basis of this, we can say that "It is raining" occupies the same formal role in English as "Es regnet" does in German and "Det regnar" in Swedish. To proto-assert that it is raining - note the oratio obliqua - it is not necessary to utter "It is raining" (and to be disposed to follow it up with other appropriate English constructions): any utterance formally analogous will do.

It is proto-asserting (or thinking-out-loud) that p in this sense that the analogy theory wants to derive thinking that p from: thinking that p is conceived as "a silent move that is a formal analogue" ([11], p. 186) of the proto-assertion that p. When one thinks (silently) that it is raining, a covert episode, with a content that plays (more or less) the same formal role as "It is raining" in English, occurs to one, and one is prepared to follow it up with other episodes having a content formally analogous to the constructions which rules of English require as an accompaniment to "It is raining" ([11], e.g. p. 184). Thinking is seen as an inner manifestation of habits of response that have their primary manifestation in overt speech (ibid., p. 189). It should be observed that thinking that p, so conceived, does not entail that any specific sentence, materially characterized, occurs to the thinker.

The crucial question, in assessing the analogy theory, is whether the construct of proto-assertion or thinking-out-loud really satisfies both (i) and (ii) above. I think Aune's extensional description of proto-assertion fulfils the second requirement. But will an inner episode modelled on proto-asserting that p, so understood, really amount to something approaching thinking that p? I think not.

The verbal behaviour appealed to in the explication of proto-assertion must be behaviour in a very attenuated sense, since reference to inner episodes is to be avoided: it must be construed as bodily reflexes, bodily movements elicited by external stimuli, as comparable to e.g. the patellar-reflex or the shivering with cold. But surely semantical terms are not applicable to such behaviour; it cannot be meaningful in the requisite sense. At least Sellars seems aware of this ([47], pp. 523, 527, 530), and his insistence that meaningful verbal behaviour must be governed by normative rules can be viewed as a

move designed to block this criticism. This move, however, merely relocates the difficulty: does it make sense to say that normative rules are in force and regulate behaviour in a world where "behaviour" carries only this attenuated sense? Does not the concept of a norm presuppose the concept of intentional action? (Cf. my definition of a normative rule as rule that it is possible intentionally to violate, 3.2.2.)

In [183] Sellars tries to meet this point by a distinction between two kinds of norms: "ought-to-do" rules and "ought-to-be" rules (pp. 506-13). The first construction is indeed, Sellars agrees, only predicable of subjects who are capable of thinking (but not necessarily in the inner sense, Sellars insists ([187], pp. 120-1)) and acting at will. But the latter can be predicated of subjects less generously endowed, indeed even of inanimate things: one can say e.g. "The glasses ought to be clean". Consequently, the statement that reality conforms to this latter kind of rule does not entail that the subject (e.g. the glasses) exercised any conceptual ability. To be meaningful, Sellars maintains, it is sufficient that the utterance of some sounds is in accordance with some rule in this second sense. It is only at later stages in the process of language-learning that a person becomes capable of formulating "ought-to-do" rules - like "One ought to bring it about that people be such that they utter 'This is red' when in front of red objects" - and of acting in compliance with such a rule. One is then fit to become a teacher of language oneself ([183], pp. 512-3).

To this it might justifiably be objected first that the fact that verbal responses are in regular conformity with an ought-to-be rule does not ensure their meaningfulness. The parrot Polly can be such that the rule "Polly ought to be such that she utters 'This is red' whenever she is presented with something red" is satisfied, yet she does not meaningfully say "This is red".

There is also another hitch. As Sellars himself notes ([183], p. 508), ought-to-be's imply ought-to-do's (given that we exclude the "ought" of probability): e.g. the norm "People ought to be such that they utter sound S in circumstances C" can be in force only if a norm of the form "x ought to bring it about that people be such that they utter S in C" is in force. But Sellars has conceded that the applicability of the latter type of rule presupposes conceptuality - to be sure, not necessarily in the form of inner mental episodes, but at least in that of thinkings-out-loud. However, there is an acute problem about how to construe this presupposed conceptuality, these thinkings-out-loud. Are they definable in a behaviouristic way, in terms of uniformities of observable behaviour? In that case, the appeal to prescriptive rules turns out to be a sham manoeuvre, a superfluous detour towards a strictly

behaviouristic conception of conceptuality - a position which Sellars sensibly seems to repudiate explicitly, as noted above. Or does Sellars stick to his doctrine that a reference to prescriptive rules (ought-to-be's and ought-to-do's) is indispensable in the analysis of the presupposed conceptuality? If so, he is caught in an infinite regress.²⁵

It is futile to try to meet this objection by denying the dependence of ought-to-be's on ought-to-do's or by reformulating the linguistic rules in terms of the "may" of permission instead of the "ought" of obligation,²⁶ for the argument can be formulated more generally as follows. Episodes of thinking are characterizable in semantical terms. Hence, whatever is to serve as the model of thinking must be characterizable in these terms as well. As analogy theorists seem to agree, verbal behaviour, conceived in a strictly behaviouristic fashion, is not so characterizable. Thus, the suggestion that a reference to prescriptive notions such as that of an obligation or a permission be added: the model is constituted by behaviour falling under such notions. To this the objection is that these prescriptive notions presuppose the concept of thought in the shape of intentional behaviour: only in a world where there is intentional behaviour does it make sense to say that something is obliged or permitted to do something.

I conclude, then, that Sellars and Aune have not succeeded in laying bare a form of verbal behaviour that fulfils both (i) and (ii). The thrust of the argument above is that (i) and (ii) can scarcely be jointly satisfied: either the (purportedly) linguistic behaviour violates (i) in that it cannot be sensibly characterized in semantical terms, or it upsets (ii) because it presupposes thought (in the inner sense). The analogy theorists have not made good their claim that a situation could have existed where we had "no conception of an inner mental occurrence but were entirely at home with the notion of intelligent speech" ([11], p. 183).

I hold the reverse of the Sellars-Aune view, namely that semantical terms ("meaning", "reference" etc.) have their primary application to thought episodes and that overt speech becomes characterizable in terms of them only to the extent that it becomes connected with thought; in other words, that utterances have meaning only because people regularly mean something by them.²⁷ Since there are meaningful utterances that are habitual, one cannot claim that an utterance has meaning only on those occasions when the speaker means - i.e. wants to convey - something by his words. But habitual verbal responses can be meaningful only when they have grown out of intentional verbal practice.

3. The analogy theory represents the inner process of thinking not only as

dependent on linguistic ability, but also as parasitic on a species of public phenomena, namely utterances. The reasoning above is intended to undermine only the second aspect. Let me now argue for the language-independence of thinking in the shape of (T8) and (T9). I shall put forward three arguments for these two theses, the first of which supports only (T8).

(1) As has been pointed out ([110], p. 236), one can think things that one is not capable of expressing fully in words. I know how a certain tune goes, but I am unable to describe it in words; all I can say is "It goes like this ..." and sing it or play it on some instrument. It is reasonable to assume that I can only sing or play a tune (without notes) if I have some mental representation of it. In the case at hand, it seems most natural to assume that this representation consists in some sound-image derived from a hearing of the tune. (I shall return to representation in images in a moment.)

(2) A second argument is provided by the obvious fact that thoughts and beliefs (together with wants) are frequently ascribed to animals and preverbal children in ordinary discourse. Aune attempts to undermine this by arguing that

"the mere fact that thinking is commonly ascribed to animals ... is no guarantee that it is truly ascribed to them, or that such ascriptions are not in any way confused" ([11], p. 219).

There is clearly something to this retort: there is nothing sacrosanct about ordinary language; it may stand in need of correction in the light of philosophical insight. All the same, it is a reply that must be handled with care - or one's philosophical theories easily become irrefutable. If one sets out, as Aune does, to clarify the common-sensical notion of thinking, the non-technical usage of "to think" and related words provide the hard data against which the theory has to be checked. But if one then allows oneself to criticize as confused the usages that do not accord with the theory, there is an evident risk of placing one's theory beyond falsification. To eliminate this risk, we must demand arguments independent of the proposed theory for the ruling out of the usages in question.

This demand must be made with a particular emphasis in the case at hand, since there are many striking parallels between animal behaviour and human thought-guided action that tempt one almost irresistibly towards the ascriptions in question. For example, if a monkey can be trained to perform a trick for the reward of coins, which it exchanges for fruit by putting them in a slot-machine - and even to hoard coins as it were for future use - then, indeed, great determination is required for one to withhold ascription of thought to the monkey. In view of considerations like these, the fact that Aune, so far as I can see, puts forward no independent argument against these ascriptions creates the impression that his theory has concentrated on sophist-

icated thinking to the exclusion of less advanced forms.

Perhaps, however, there is such an argument after all. A belief must have propositional content. Some philosophers ([6], pp. 25-7; [214], pp. 172-3) have, in effect, disputed this dictum by holding that beliefs, in the sense attributed to animals, are referentially transparent and even extensional (i.e. substitution of coextensive predicates is permissible) ([6], pp. 25-7). But the view that an animal belief is a direct relation to individual things and classes of things is quite implausible. (E.g. the joyful behaviour of a dog justifies attribution to it of the belief that its master is approaching, but not of the belief that the man who recently maltreated it severely is approaching, although, unbeknownst to the dog, the master = the maltreater.) We should stick to the doctrine that all beliefs have propositional content, that the context "believes that ..." is intensional. But now it may be argued (cf. [194], esp. pp. 22ff) that there is such a difference between ourselves and animals that we are not even in principle capable of specifying the content of an animal belief. This, if true, should shake a confidence in the legitimacy of assignations of beliefs to animals. But I do not think it is true. Without involving myself too deeply in the issue, I want to suggest that the key lies in the fact that the concepts of animals are derived from their perceptual data; and their perceptual data can be conjectured from a comparison between the physiology of their sense-organs and ours. So, to my mind, this argument against animal beliefs fails.

(3) Considerations from the learning of conventional languages give powerful support to the idea that propositional thinking must occur in another medium than that of a conventional or learnt language. To learn the meaning of a certain type of speech-sound it is necessary to single out certain aspects from the wealth of perceptual data, namely a speech-sound token and what it is supposed to designate, e.g. an instance of a colour. It is further necessary to recognize aspects of later experiences as being of the same type as these. Lastly, it is necessary to connect in some fashion the type of speech-sound with the other type of feature of one's perceptual field. All the underlined words certainly mark exercises of conceptual ability, i.e. thinking. The process can be summed up as follows:

(L) To learn the meaning of a type of speech-sound S, it is necessary (for the learner) to formulate in thought a regularity that is expressible in a rule like "Sounds of type S are uttered (by parents etc.) in and only in circumstances of the type C".

But in what medium does the learner represent this rule to himself? In some cases, it may be in the symbols of some language he has learnt. But, on penal-

ty of an infinite regress, this cannot always be the case: the meaning-rules of the first sounds of the first conventional language he acquires must be represented, thought, in a symbol-system the significance of which he has not learnt. So, there is a natural or unlearnt medium of thinking-that (cf. [80], pp. 59-64).

In [178] Sellars attempts to show that something like (L) need not be postulated to account for the learning of conventional language. Sellars wants (pp. 324-7) to draw a parallel between an individual's learning of language and a species' acquisition of patterns of instinctive or reflexive behaviour in the course of evolution. At some point in the course of evolutionary development, a teleological explanation of some fact concerning a species - e.g. that the fur of a weasel turns white in the winter - becomes applicable. But the reference to the survival and well-being of the species in such an explanation does not imply that the members of the species have any conception of these. Similarly, Sellars suggests, to say that some children, at some stage in their training, begin to utter some speech-sounds in certain circumstances "because of the (ought-to-be) rules of language" does not entail that the children have any conception of these rules. The explanation only requires that these habits of verbal response are brought about by other people who obey - and hence have a conception of - the implied ought-to-do's: "Trainees conform to ought-to-be's because trainers obey they corresponding ought-to-do's ([187], p. 123).

But the fact that the (physical) behaviour of something is made to conform to ought-to-be's by agents obeying ought-to-do's does not guarantee that the behaviour is meaningful. Suppose the strokes of a clock correspond to the time, because it is constructed by people animated by the rule that they ought to produce clocks which conform to "Clocks ought to be such that their strokes correspond to the time". Then this, though enough to justify the (rather coarse) explanation that the clock behaves as it does because it is as it ought, does not imply that the clock means anything by its strokes. So, if language-learning had followed this pattern, it would not, to be sure, presuppose any mental representations on the part of the trainees, but its end-product would not be meaningful linguistic behaviour. On this hypothesis, language-learning is not learning in the ordinary sense at all.

Perhaps, however, this is a misinterpretation of Sellars. In a couple of passages (e.g. [178], pp. 327, 333) Sellars talks of the learning of language as conditioning. However, he does not attempt to clarify what is meant by "conditioning" - something which is surely necessary, for it is arguable that conditioning presupposes thought. This appears to be so already in animal

cases of classical conditioning: a certain sound-token makes a dog salivate because the dog recognizes it as being of a certain type and remembers that he has been given food when this type of sound has occurred. It appears that conditioning is only possible where there is some conceptual ability, for it is types of situations that are associated. But even if this is not generally true, it remains to be shown that it is not true of the conditioning involved in learning language, i.e. not true in a case where one learns to use something with meaning, where one learns to behave meaningfully and intelligently.

These interpretations put Sellars in a dilemma: if his model of language-learning is so interpreted that it steers clear of assuming pre-verbal thinking, it cannot make it comprehensible how speech-sounds can be uttered with meaning, while if it is given a form in which it stands any chance of making linguistic behaviour meaningful, it is a moot-point whether it succeeds in eschewing the assumption of pre-verbal thinking.

4. So far as I can see, there is no escape from the conclusion that there is thinking-that in a representational system that is natural in the sense that the meanings of its symbols are not learnt - hence, in a symbol-system that must be altogether different from those of conventional languages. But it is easy to understand the urge to evade this conclusion, for it raises a very recalcitrant problem: what is this natural medium of thought? It is not that no answer readily suggests itself, but rather that what suggests itself is so disreputable that one mentions it with trepidation: the classic theory that the natural medium of thought is that of images. Since the term "image" exclusively suggests replicas of visual sense-impression, and there is no reason to exclude reproductions of other sense-impressions, let us replace it by "isomorphic symbol". Let me briefly dwell on the merits and demerits of this theory.

What could it mean to say that an isomorphic symbol designates or is a symbol for something? Traditionally, the relation of resemblance has been invoked to explicate this notion. But this will not do, because resemblance is a reflexive and symmetric relation, while designation is irreflexive and asymmetric. Perhaps matters can be improved by relying on causality, which possesses the requisite formal properties. Normally, the capacity to produce isomorphic symbols of a certain sort is acquired by perceiving property-instantiations of the relevant kind: the capacity to evoke images of red is acquired by having visual impressions of red etc. So, to make a concrete proposal, suppose this is introduced as a necessary requirement along with a resemblance-condition: x can be an isomorphic symbol of F-ness only if x belongs to a class of

entities the original members of which were directly caused by perceptual instances of F. The objection that it is possible imaginatively to construct e.g. an image of a particular shade of red from perception of neighbouring shades could be met by weakening the demand to "a class of entities the original members of which were directly caused by perceptual instances of the properties which they resemble". In the case of isomorphic symbols of individuals, however, it seems that such a loosening of the causal demand is not necessary: a causal link, even if very indirect, to the individual is a sine qua non because one individual can be a duplicate of another.

Even these sketchy remarks are sufficient to enable one to see how an isomorphic symbol can function as a symbol without its significance ever having been learnt. Since an isomorphic symbol of F-ness is by definition constituted by that which sense-impressions of instances of F give rise to in one's thought, it could not possibly be correlated with that of which it is a symbol by a non-trivial convention. To set up a convention associating an instance of F with an isomorphic symbol of F-ness, one must identify instances of F in some way. But ultimately this identification can only be done by using isomorphic symbols of F-ness in a perceptual identification; hence the resulting convention will be completely trivial, comparable to "By 'red' is to be meant red". Such a rule cannot be the object of learning. What makes something an isomorphic symbol of F-ness is not a convention, but a natural fact: roughly, that it resembles perceptual instances of F more than other instances and belongs to a class which originated in such perceptual contents.

Graver difficulties, however, lie ahead. The arguments from language-learning and from animal beliefs require that the natural code be rich enough to express propositions, but how is a proposition to be formulated in a system of isomorphic symbols? The trouble is that such a system cannot, it seems, have much of syntax. Nevertheless, it is not manifestly absurd to think that elementary syntactical structures can be constructed out of relations that unquestionably hold between isomorphic symbols, e.g. relations of temporal and causal ordering and incompatibility. For instance, the thought that something in one's perceptual field is F could be understood as the perception of this item causing the application of an image of F-ness to it (whatever that may exactly mean). And thinking of the same thing, in its perceptual absence, that it is F could be viewed as an image of the individual thing causing an image of F-ness to arise immediately after it, assuming that the images are compatible. Furthermore, a person's thinking the conjunction of such propositions, p and q, may be his thinking that p and his thinking that q in immediate succession, where neither p and q, nor p and -q, are explicitly incompatible.

The disjunctive thought that p or q is perhaps nothing but switching back and forth from the thought that p to the thought that q, where p and q explicitly exclude each other. (Merely switching once from p to q, where incompatibility holds, is not enough, since this could be a change of mind.) Finally, when the thought that p causes the thought that q and p (perhaps in conjunction with other thoughts of the person) is incompatible with ¬q, we may have a situation which could be interpreted as a rudimentary form of thinking the conditional thought that if (or since) p then q. These brief remarks are, I think, enough to show that it is not obvious - as it may appear at first glance - that propositions cannot find expression in a code of isomorphic symbols. But can propositions sufficiently sophisticated for the learning of conventional languages be formulated in such a code?

What degree of sophistication is needed? Fodor has persuasively argued ([80], pp. 79-83) that the natural code must have at least the same expressive powers (in respect of what is true-or-false) as the conventional language to be learnt. If this is correct, it seems certain that the code of isomorphic symbols does not fit the bill, for how could the syntactical complexity and the abstract concepts of civilized conventional languages possibly be captured by the isomorphic code? To be sure, Fodor adds (pp. 84-5) a mitigating consideration: the natural code-counterparts of some manageable conventional constructions may be so complex that the powers of memory and imagination are in fact insufficient to hold them simultaneously before the mind. For, as I have noted, it is not necessary that a learner represents to himself every meaning-rule in the natural code; once he has acquired a certain portion of a conventional language, he can formulate meaning-rules for the remaining segments in the vocabulary of the ever-expanding learnt part. I shall not attempt to judge whether this qualification is enough to save the idea that the natural code presupposed in the learning of language is that of isomorphic symbols. A stand on this difficult matter is fortunately not essential for my purposes.

To sum up the argument of this section. Reasons have been found to assent to both (T8) and (T9). In the case of (T8), I have also been able to give an account of these non-conventional constituents of propositional content: they are isomorphic symbols.²⁸ But it is doubtful whether the same answer works in the case of (T9), doubtful whether the code of isomorphic symbols is resourceful enough to express propositions, at least to the extent demanded by the learning-argument (whatever that extent precisely is). Further, since it is disputable also whether this natural code is rich enough to formulate conditionals, we cannot either confidently affirm that the domain of applicability

of reasons for action extend beyond language-users. The problem of the applicability of the reason-terminology, however, is not peculiar to my theory of reasons for action. For this theory it is only important that an affirmative answer to (T8) has been given in its essentials. If it should prove impossible to work out an isomorphic code capable of expressing propositions, individual and general images must be thought of as occurring in syntactical frames of another code when they are elements of propositions.²⁹

I hope that the theses (T1)-(T3) and (T6)-(T8) together with the comments on them make the concept of thinking-that sufficiently clear for my purposes.

4. CONDITIONALS AND POSSIBILITIES

4.1 An Analysis of Conditionals.

1. In chap. 1 I made the claim that reasons for action are propositions that are expressible in "if" and "only if" sentences. Unfortunately, it is far from clear how these sentences are to be interpreted - provided, of course, that they are not taken to be expressions of material or strict implications, which would be too wide or too narrow, respectively, for my purposes. Therefore, I now undertake to clarify what I intend these constructions to convey. It is enough for the central argument of this work if the definitions I am about to offer are accepted as (I hope perspicuous) stipulations on my part. But I do believe that they are good approximations to ordinary usage - that is the reason why I have chosen to formulate my concepts by means of the non-technical terms "if" and "only if".

Let me start by revealing how I would like to construe what Mackie calls ([127], chap. 3) an "open" conditional, i.e. a conditional whose linguistic garb leaves it open whether or not the antecedent is/was/will be actualized. After having explained this definition, I will briefly consider counterfactuals, and then later on open "only if" sentences.

Let "M" represent logical possibility and let "s" designate the content of the system of convictions of the speaker. In the case of ordinary mortals, there will invariably be lacunae in this system, matters about which the speaker is totally in the dark. So there is such a thing as an expansion or saturation of s, i.e. supplementing s with further propositions that are taken to be true. I now define:

(C1) If $\underline{p}, \underline{q} \Leftrightarrow \underline{M}(\underline{s} \ \& \ \underline{p}) \ \& \ \underline{M}(\underline{s} \ \& \ \underline{-q}) \ \& \ \exists(\text{an expansion of } \underline{s}, \underline{s}^*)(\underline{M}(\underline{s}^* \ \& \ \underline{p}) \ \& \ \underline{M}(\underline{s}^* \ \& \ \underline{-q}))$.¹

According to (C1), "If $\underline{p}, \underline{q}$ " asserts that, though $\underline{p}$ and $\underline{-q}$ individually are consistent with the speaker's convictions, there is an expansion of these convictions which, while retaining these compatibilities, is inconsistent with $\underline{p} \ \& \ \underline{-q}$, i.e. $\underline{q}$ is deducible from this extension enriched with $\underline{p}$. The idea that someone who puts forward a conditional is stating something about the deducibility of the consequent from the antecedent, together with implicit assumptions, is of course not new; Roderick Chisholm has experimented with it as early as in [46]. But, as will become apparent, my analysis diverges from his on some vital points.

To make my proposal more precise, it is necessary to distinguish between two major senses of the phrase "it is possible that $\underline{p}$ ", namely absolute (logical) and relative possibility. Something is absolutely possible iff it

involves no contradiction. In relative possibility p is related to something that is implicit and declared to be (logically) possible in conjunction with it. Relative possibility can be transformed into absolute possibility if the implicit assumptions are made explicit and conjugated with the explicit proposition. This is what I have done in the analyses of (C1): of course, somebody who says that if p then q does not explicitly refer to his convictions: he uses the concept of relative (im)possibility. But in the interest of clarity, I have made these background-assumptions explicit and thus transformed the relative (im)possibility into an absolute one. This is enough to show that (C1) cannot pretend to present a synonymy, but has to rest content with capturing the truth-conditions of the definiendum.

Within the class of relative possibility one can distinguish several subclasses, depending on precisely to what the possibility is supposed to be relative. First, p may be possible relative to the system of all true propositions; in this case, I shall speak of ontic possibility.² In a fully deterministic world, p is ontically possible iff p is/was/will be an actual fact of the world. For a false proposition will here inevitably conflict with the conjunction of some causal law and statements asserting its initial conditions.

Secondly, the possibility of p may be relative to what is held true by a particular speaker at a particular time. Here I will talk about (subjective) epistemic possibility. (Of course, there are also many distinguishable concepts of intersubjective epistemic possibility, of possibility relative to what is thought true by various sets of people.)

Both of these concepts of relative possibility enter into the analysis of hypotheticals: epistemic possibility into the first two conjuncts and ontic possibility into the last, existentially quantified conjunct. I think it fairly obvious that it would not do to utilize the same concept twice over. We could hardly use ontic possibility in the first two conjuncts, for if the world is fully deterministic - surely a supposition we would not want to consider incompatible with the employment of conditionals - it would follow from the two first conjuncts that p and $\neg q$ were both ontically necessary; hence, that it would be false that their conjunction is ontically impossible. Besides, we frequently say things like "If p , then ..., but if $\neg p$, then ...". On the other hand, it is evident that we could not replace the existentially quantified limb with " $\neg M(s \ \& \ p \ \& \ \neg q)$ ", for then a person with a sufficiently restricted set of convictions could almost always deny the truth of the conjunct.

I have spoken of the convictions of the speaker. By this I mean propositions whose truth the speaker holds for certain. So if the speaker regards r (or the

existence of a proposition with the formal properties of r) as merely probable, r is not included in s or its extension, s*. However, mere probabilities (or epistemic possibilities) are not irrelevant to the act of putting forward conditionals; they form a legitimate part of the background, if the conditional propounded is epistemically guarded, i.e. if "it is probable (epistemically possible) that ..." is tacked onto the conditional.

Sometimes, however, no background-convictions are necessary for generating the impossibility of the last conjunct of the existentially quantified limb of (C1). In this case, q is an explicit logical consequence of p. But it is more common that the speaker draws upon some implicit convictions of his. Even the case where p implicitly entails q can be viewed as a case of this sort; here the speaker can be seen as implicitly relying on those analytical truths that make the entailment apparent.

An especially important type of case where background-assumptions are relied upon is when q is asserted to be a causal consequence of p. Here the speaker must have some causal law in mind that covers the facts that p and q. It might be thought that this generates a vicious infinite regress, for is not a causal law itself a conditional of the causal kind? But then the speaker needs to have access to another causal law which backs up the first and so on. This line of thought is, however, mistaken. As e.g. Donald Davidson (e.g. in [60], p. 92 and [62], pp. 91-2) has emphasized, in order to put forward a singular causal judgement one need not be able to state any law which the particular connection instances; it is enough if one is justified in holding that there is such a law. I think this point can be adapted to general causal judgements as well. If so, the threat of regress is countered, for - in contrast to the project of formulating an infinite number of laws - there is nothing viciously regressive about maintaining that the hierarchy of causal laws is infinite, that for every causal law there is another from which it can be derived. It might be objected, however, that according to this account of the matter, causal judgements become excessively hard to verify, since we cannot even in principle establish the existence of an endless hierarchy of laws. To this objection I have two replies. First, my theses do not entail this view of causation, because I have not here assented to one premiss necessary for it: that causal laws are expressible in conditionals as I understand them. But secondly, I do not think this skeptical conclusion concerning the verification of causal statements should deter one from assenting to it. I am inclined to believe that the verification of causal judgements (concerning outer objects; for this qualification, see 5.4.4) is such a precarious matter, though I would not deny that ordinary experimental

methods can justify a rather high degree of assurance of their truth or falsity.

It seems to me that the conditional relationship, as delineated in (C1), adequately captures every relationship between propositions that is of interest in the treatment of reasons for action. I have tried to show that it is capable of dealing with causal and logical consequences, but it can also handle conventional "effects". Suppose that "p" represents the proposition that my arm goes up and "q" that my arm moves in a greeting gesture; then the conditional "If I bring about p, then q" can be supported by a convention regulating which bodily movements are to be counted as greetings. In virtue of the fact that the greeting movement is a conventional effect of the rising of the arm, I bring about the former by bringing about the latter.

However, (C1) faces, as it stands, a grave difficulty: surely, the negation of "If p, q" is not the negation of the whole conjunction constituting the right-hand side of the equivalence; " $\neg(\text{if } p, q)$ " is not made true just because e.g. p is not possible relative to the system of accessible truths, s. But exactly how is " $\neg(\text{if } p, q)$ " to be understood? Robert C. Stalnaker has suggested ([193], p. 173) that this construction is equivalent to "if p, $\neg q$ ", i.e. that the denial of a conditional is the denial of its consequent. However, an example of Ryle's ([172], pp. 127-8) shows that this is unacceptable: the negation of "If I walk under that ladder, I shall meet trouble during the day" is certainly not "If I walk under that ladder, I shall not meet trouble during the day"; for that would be equally superstitious: the first saying that walking under that ladder would lead me into trouble, the other saying that it protects me from getting into trouble. Similarly, the negation of "If one jogs a mile every morning, one will stay healthy all one's life" is evidently not "If one jogs a mile every morning, one will not stay healthy all one's life". Generally put, denying a conditional cannot be the same as denying its consequent when the antecedent is affirmed, for the original conditional is false also if the consequent is fulfilled together with the antecedent, but the former is true because of factors having nothing to do with the latter. As an interpretation of the negated conditional, I would instead opt for what is in effect Ryle's proposal, that the relevant negation is equivalent to "It is possible that I walk under that ladder and still do not meet trouble during the day", more precisely, that it is only the existentially quantified conjunct of (C1) that is being negated.

But how can it come about that only this last conjunct is negated? The answer is that the first two conjuncts - $\underline{M}(\underline{s} \ \& \ p)$ & $\underline{M}(\underline{s} \ \& \ \neg q)$ - are pre-supposed rather than asserted in issuing the conditional. In putting forward

the conditional one supposes that p . The speech-act of supposing is like that of referring in that neither, unlike e.g. the act of assertion, can stand on its own. They both cry out for completion: the act of reference for the act of predicating and the act of supposal for the drawing of consequences from it. (This is subject to a qualification to be added later on.) There is a further parallel to be drawn between referring and supposing: just as the existence of what is referred to is presupposed in the act of reference, so the (epistemic) possibility of the proposition supposed is presupposed in the act of supposing it. Thus I maintain that the reason why $M(\underline{s} \ \& \ p)$ is not negated in the negation of "If p , q ", is the same as the reason why the existence of the King of France is not negated when "The King of France is bald" is negated. (I would also suggest that just as there is a problem about the truth-value of a statement when the existence-presupposition of its referring-phrase is not satisfied, so there is a problem about the truth-value of a conditional when the possibility-presupposition of its protasis is not satisfied.)

Furthermore, when one supposes p in order to derive q it is presupposed that p has a role to play in this derivation, that the background assumptions alone are not sufficient for the inference of q , that is, that $\neg q$ is compatible with the background-truths by themselves.

(Notice that if it is $M(\underline{s} \ \& \ p)$ and $M(\underline{s} \ \& \ \neg q)$, there will in practice always be expansions of $\underline{s}$ that preserve these relations - there are lots of true propositions that are irrelevant to p and $\neg q$. So the crucial part of the existentially quantified component is the last conjunct: it can safely be assumed that if a conditional is negated, it is because no expansion answers to this conjunct.)

The upshot of these remarks is that there is another reason - other than the one about the transformation of relative possibility into absolute - why (C1) can only be taken as determining the truth-conditions of "if p , q ". In this way I dodge a criticism Mackie directs at Chisholm:

"Someone who uses a conditional may indeed be relying on other premisses; but if so he is not saying that they are cotenable with the supposition, but merely in fact retaining them for use along with the supposition" ([127], p. 69).

What about $M(\underline{s}^* \ \& \ \neg p)$ and $M(\underline{s}^* \ \& \ q)$ - and a fortiori $M(\underline{s} \ \& \ \neg p)$ and $M(\underline{s} \ \& \ q)$? They are both entailed by the analysans of (C1). For $\forall$ (expansions $\underline{s}^*) \ L(M(\underline{s}^* \ \& \ \neg q) \ \& \ \neg M(\underline{s}^* \ \& \ p \ \& \ \neg q) \Rightarrow M(\underline{s}^* \ \& \ \neg p \ \& \ \neg q))$; and $\forall(\underline{s}^*) \ L(M(\underline{s}^* \ \& \ \neg p \ \& \ \neg q) \Rightarrow M(\underline{s}^* \ \& \ \neg p))$. Similarly, $\forall(\underline{s}^*) \ L(M(\underline{s}^* \ \& \ p) \ \& \ \neg M(\underline{s}^* \ \& \ p \ \& \ \neg q) \Rightarrow M(\underline{s}^* \ \& \ p \ \& \ q))$.

2. It is notable that, on my account, "If $\underline{p}$, $\underline{q}$ " is incompatible with "If $\neg \underline{p}$, $\underline{q}$ ". For if for one expansion $\underline{s}^*$ it is $\neg M(\underline{s}^* \ \& \ \underline{p} \ \& \ \neg \underline{q})$, it can be $\neg M(\underline{s}^* \ \& \ \neg \underline{p} \ \& \ \neg \underline{q})$ only if $\neg \underline{q}$ is incompatible with $\underline{s}^*$ *per se*, but this is excluded by $M(\underline{s}^* \ \& \ \neg \underline{q})$. Acknowledging this incompatibility puts me in direct opposition to Stalnaker ([193], pp. 167-8) and Mackie ([127], p. 109). This divergence of opinion has wider significance, because, as Mackie has pointed out (*ibid.*), this more or less commits one to a certain view concerning the validity of the rule of contraposition. For it seems certain that "If $\neg \underline{q}$, $\neg \underline{p}$ " and "If $\neg \underline{q}$ $\underline{p}$ " - i.e. the contraposed versions of the original conditionals - contradict each other. Consequently, one seems compelled either to deny the validity of contraposition or proclaim "If $\underline{p}$, $\underline{q}$ " and "If $\neg \underline{p}$, $\underline{q}$ " to be incompatible. Stalnaker (pp. 173-4) and Mackie choose the former (though Mackie restricts this to the direct, as opposed to the *reductio ad absurdum*, use of conditionals) whereas I opt for the latter alternative. This accords better with my intuitions, but not only that: one cannot claim that "If I bring about $\underline{p}$, then $\underline{q}$ " gives a reason for bringing about $\underline{p}$ if one holds it to be compatible with "If I do not bring about $\underline{p}$, $\underline{q}$ "; for whatever is a reason for causing $\underline{p}$ must to that extent be a reason against not bringing it about.

Mackie is prepared to set up ([127], pp. 110-1) special rules for the use of conditional constructions in *reductio*-contexts. Here contraposition is said to be a legitimate form of manipulation, and formulae like "If $\underline{p}$, $\underline{q}$ " and "If $\underline{p}$, $\neg \underline{q}$ " to be compatible. This is a concession to those who are sympathetically inclined to the idea that "if-then" can be analyzed in terms of material implication, since on this interpretation these formulae will always come out as consistent.³ But can I myself avoid introducing a special *reductio*-employment? It might seem that my own interpretation will not work in *reductio*-contexts because the speaker here does not believe in the possibility of the antecedent and, consequently, can hardly be willing to presuppose it.

Belief that $\underline{p}$ is compatible with the system of truths accessible to a speaker is, however, not a necessary prerequisite for him being entitled to presuppose the compatibility. It is enough that his audience believe in this compatibility, for there is such a thing as to put oneself in the place of others and to argue from their assumptions. A *reductio* is an attempt to reduce someone *else's* belief in (the possibility of) a proposition to absurdity by adding the proposition to the stock of convictions that are *shared* by speaker and hearer alike and then demonstrating that, with legitimate extensions of this stock, a contradiction becomes inferable, e.g. that for some expansion $\underline{s}^*$ both $\neg M(\underline{s}^* \ \& \ \underline{p} \ \& \ \neg \underline{q})$ and $\neg M(\underline{s}^* \ \& \ \underline{p} \ \& \ \underline{q})$. The conclusion is that the belief that $\underline{p}$ could consistently be added to the stock of truths is false. This

procedure does not require that the speaker believes that his premisses of the form "If $\underline{p}$, $\underline{q}$ " and "If $\underline{p}$, $\neg \underline{q}$ " are true, still less that they are in fact true. In fact, neither of them is - as the speaker recognizes - true, since it is not possible that $\underline{p}$. What is required for the argument to have a point is that there be someone who assumes things that permit him to suppose $\underline{p}$ and hence compel him to embrace both conditionals. The argument reaches its end if the addressee admits that $\underline{p}$ is not possible and, consequently, like the speaker himself, rejects both of the conditionals. But of course the hearer may also react in a way the speaker does not anticipate: by giving up one of the shared convictions that the argument rests upon. According to this account, the same rules hold for conditionals in *reductio*-contexts: contraposition is valid, "If $\underline{p}$, $\underline{q}$ " and "If $\underline{p}$, $\neg \underline{q}$ " are incompatible (and neither of them true), and "If $\underline{q}$, $\neg \underline{p}$ " and "If $\neg \underline{q}$, $\neg \underline{p}$ " are incompatible.

There is another well-known principle about which Stalnaker and I disagree, namely the principle of transitivity. Stalnaker argues ([193], p. 173) that if this principle holds good, then one can legitimately infer "If $\underline{p} \ \& \ \underline{r}$, $\underline{q}$ " from "If $\underline{p}$, then $\underline{q}$ " (the suppressed additional premiss is "If $\underline{p} \ \& \ \underline{r}$, $\underline{p}$ "). But, he continues, an example shows that this is not so: from "If this match were struck, it would light" one cannot infer "If this match were soaked in water overnight and it were struck, it would light". To this I wish to reply that transitivity obtains so long as the set of background assumptions remains constant. $\underline{p} \ \& \ \underline{r}$ must be consistent with the background convictions used in the derivation of $\underline{q}$ from $\underline{p}$. Stalnaker's illustration violates this restriction: the subjunctive conditional "If this match were struck, it would light" is only acceptable if the background assumptions include the proposition that the match is dry, but to this set the proposition that the match was soaked in water overnight cannot be consistently added.

How do counterfactuals fit into the picture? It is quite clear that (C1) does not apply without qualification to counterfactuals, because they are formulated in such a way that the speaker signals that the negation of the antecedent belongs to his stock of convictions. Hence, to maintain consistency when the counterfactual supposition is added to the stock, its direct negation at least must be discarded. Further, if the system contains propositions that entail the negation of the counterfactual antecedent these must be abandoned too. But this does not suffice. Consider a problem that exercises Chisholm ([46], pp. 299ff). If a speaker is convinced that $\underline{p}$, he may justifiably conclude that either $\underline{p}$ or $\underline{q}$. Suppose now that $\neg \underline{p}$ is counterfactually assumed and that only $\underline{p}$ is deleted; then the speaker has in his possession premises that entail $\underline{q}$, i.e. he has a right to say that if $\neg \underline{p}$ had obtained then $\underline{q}$ would have

obtained too. This is surely unacceptable. A remedy that lies close at hand is to insist that the speaker abandons together with p every proposition that he has used p to establish, every proposition that for him is dependent upon p. In terms of our example, this means that the speaker must give up the disjunction p or q along with p, since we have assumed that he used the latter to prove the former.

But it seems that these considerations are not sufficient to handle all problems of deletion. Why are we, in Goodman's famous example ([90], pp. 20-1), prone to retain the fact that the match was dry rather than the fact that it did not light, when we counterfactually assume that it was struck, i.e. prone to assert "If it had been struck, it would have lit" rather than "If it had been struck, it would not have been dry"? It might be suggested that a match's lighting or not, but not its being dry or not, is known to be dependent on whether it is being struck or not. But the "dependency" is here causal dependency. Consequently, this proposal will be non-circular only if causation does not involve counterfactuality. So, it can be seen that the problems of deletion lead us into very deep waters indeed. Fortunately, I need not take a stand on the issue of deletion, since counterfactuals will not play any leading part in my theory.

3. It cannot be denied that there are examples even of open "if"-sentences that (C1) does not fit. For instance, there is Austin's famous "There are biscuits on the sideboard if you want them" ([16], p. 210). Obviously, it is not here being stated that, given background assumptions, your wanting biscuits and there not being any on the sideboard are incompatible. For further counter-instances, we may think of hypothetical imperatives (e.g. "If you are humiliated, avenge yourself!") and hypothetical declarations of intention (e.g. "If I am humiliated, I shall avenge myself"); in these cases, (C1) will not even make sense.

Mackie has offered an interpretation of conditionals that easily takes care of the Austinian conditional. According to him, "to say 'If P, Q' is to assert Q within the scope of the supposition that P" ([127], p. 93). The outstanding difference between Mackie's analysis and mine is that whereas I let the assertion-operator include within its scope the whole conditional, this operator, according to Mackie, only covers the consequent; in other words, he "abandons the claim that conditionals are in the strict sense statements ... that they are in general simply true or false" (p. 93). It is precisely for this reason that Mackie is able to deal smoothly with the Austinian conditional, for here the assertion-operator evidently does not

operate on the whole conditional. And it is also for this reason that he is capable of giving a natural treatment of hypothetical imperatives and hypothetical declarations of intention; all that needs to be done is to substitute "to issue an imperative (intention-declaration) within the scope of a supposition" for "to assert something within ...".

Why not accept Mackie's analysis instead, then? Because there are, as Mackie himself acknowledges (p. 102), contexts in which it will not work: an "if"-sentence can be embedded in such constructions as "a believes that ...", "it is true that ..." and "it is not the case that ...", only if it expresses a proposition. (On the other hand, note how odd it would be to embed the Austinian conditional in these contexts.) Mackie responds (pp. 103-4) to this by conceding that there is another use of "if"-sentences in which they indeed formulate propositions. He insists however that this propositional employment is logically parasitic upon the non-propositional one. So, Mackie's approach is not really more economical than mine in the usages it recognizes. The difference between us is rather that while he sees the non-propositional employment as primary and the propositional as derivative from it, I hold that the former is found only in marginal cases (so long as one deals with conditionals whose consequents have truth-value; it assumes greater significance when one turns to imperatives and declarations of intention).

I think it must be admitted that it is only in quite exceptional cases that it is not paradoxical to refuse to assign truth-value to the whole conditional. The Austinian conditional is just such an exceptional case: clearly, the truth-value of the whole conditional here boils down to that of the consequent. Mackie senses this air of paradoxicality, and is willing to compromise to reduce it: he argues (pp. 105-8) that conditionals, even in (what is for him) their non-propositional employment, can sometimes be bestowed with truth-value, namely if they are either logical or if their antecedents are fulfilled; in the latter case they are true iff the consequent is fulfilled as well. Let us for the sake of the argument grant Mackie this (though it is not concordant with what has been said above); still, the awkwardness is not fully removed. Suppose the antecedent of "If p, q" is not realized - which is always the case when a counterfactual is properly put forward - then one cannot assign truth-value to it. But how does one then determine whether "If p, -q" or "If p, q" is acceptable (given that either is), and what does "acceptable" mean here? Moreover, what is the difference in these circumstances between the speech-acts of asserting q within the scope of the supposition p and merely supposing the conjunction of p and q?

Since I still want to make some use of the notion of asserting something

within the scope of a supposition, it is incumbent on me to elucidate the notion to some degree and to clarify its relation to the concept of supposing something in order to assert something of it. Of course, the former concept does not purport that the assertion is a consequence of the supposition, as the breaking of a vase is a consequence of its falling to the floor: the man who says to his guests that there are biscuits on the sideboard, if they want some, has asserted that there are biscuits on the sideboard. The point of the addition "if you want some" is simply to convey that the speaker considers his asserting this - i.e. his committing himself to a certain speech-act - as relevant if - and presumably only if - his guests want biscuits. Thus, if it turns out that his guests do not want any biscuits, they are in no position to accuse the host of saying something that is beside the point, as they would be in a position to do had his assertion not been qualified in the described way.

Intention-declarations and imperatives of the form "If I am humiliated I shall revenge myself" or "If you are humiliated, revenge yourself!" are to be understood in a similar way: the speaker issues an intention or imperative for a certain type of possible situation, that is, he signalizes that he regards the intention-declaration or imperative issued as binding or valid (only) in a world where the antecedent has materialized. The philosophers (e.g. [137], pt. 1, chap. 2) who try to interpret these intention-declarations and imperatives as intention-declarations and imperatives with a hypothetical content, commit an error that is the direct opposite of the one I have been charging to Mackie: Mackie restricts the scope of the statement-operator to the consequent when it should include the whole conditional, while the philosophers alluded to let the intentional or imperative operator reach across the whole conditional when it should really be restricted to the consequent. Of course, there are imperatives (intentions) having a truly conditional content, e.g. "Bring this about: that the house explodes if the door opens!". But this imperative is clearly different from "If the door opens, bring it about that the house explodes!".

How does this account of asserting within the scope of a supposition fit in with the parallel earlier drawn between supposing and referring? Admittedly, the supposition is not here part and parcel of the assertion; it has to do not with the content of the assertion, but rather with the act of asserting it. But this does not ruin the parallel with referring, because the act of reference sometimes fulfils an analogous function - I am thinking of the situation where one calls somebody's name in order to attract his attention, and then asserts something he should notice (e.g. "Algernon, some rungs are missing in

the ladder"). Here the function of the name is not to specify an entity which is the subject of predication, but has to do rather with the act of putting forward a propositional content.

4. There are a couple of constructions that are closely related to "If $\underline{p}$, $\underline{q}$ ". One was mentioned right at the beginning, viz. "Even if $\underline{p}$, $\underline{q}$ ". I will not attempt to elucidate this construction, but I must confess that I am tempted to regard it as asserting, minimally, the negation of "If $\underline{p}$, $\neg \underline{q}$ ". But I shall venture a definition of "Only if $\underline{p}$, $\underline{q}$ ", since I find the phrase useful and interpret it differently than is customary in philosophical discussions:

(C2) Only if $\underline{p}$, $\underline{q} \leftrightarrow \underline{M}(\underline{s} \ \& \ \underline{p}) \ \& \ \underline{M}(\underline{s} \ \& \ \neg \underline{q}) \ \& \ \exists$ (an expansion of $\underline{s}$, $\underline{s}^*$)
 $(\underline{M}(\underline{s}^* \ \& \ \underline{p}) \ \& \ \underline{M}(\underline{s}^* \ \& \ \neg \underline{q}) \ \& \ \neg \underline{M}(\underline{s}^* \ \& \ \underline{p} \ \& \ \neg \underline{q}) \ \& \ \neg \underline{M}(\underline{s}^* \ \& \ \neg \underline{p} \ \& \ \underline{q}))$.

That is, to my mind the definiendum asserts that $\underline{p}$ is a necessary and sufficient (given background assumptions) condition for $\underline{q}$. That $\underline{p}$ is just necessary for $\underline{q}$, I would express by e.g. "If $\neg \underline{p}$, $\neg \underline{q}$ ".

4.2 The Spectrum of "Can".

1. As will be seen in the next chapters, different senses of "can" will be involved in the analyses of notions like wanting and acting at will. It is therefore necessary to survey some of the main functions which the verb "can" can serve. For at least two reasons it is fitting to undertake this task right away. First, in defining "if"-sentences I had recourse to some concepts of possibility. Secondly, according to one well-known approach, one species of uses of "can" - of central importance in the context of action - is interpreted conditionally, as meaning roughly "does if one wants (tries, chooses etc.)".

The first thing to be done is to show the need for introducing something like this latter way of using "can". This is easily done. Consider such a statement as e.g. " $\underline{a}$ can lift this stone at $\underline{t}$ ". It is quite obvious that there is a reading of this statement, according to which it means neither "It is ontically possible that $\underline{a}$ lifts this stone at $\underline{t}$ " nor "It is epistemically possible that $\underline{a}$..." (let alone "It is logically possible that $\underline{a}$..."). For the original assertion can be true although there are conditions that will lead to $\underline{a}$'s not being motivated to lift the stone at $\underline{t}$, and a speaker can put forward the statement although he believes in the presence of such motivational conditions. When used in this way, the assertion advances a claim about certain characteristics of $\underline{a}$, e.g. his strength, and his external situation, e.g. his not being tied up. The existence of the interpretation I am alluding to is evidenced by the availability of such concepts as those of ability (or capability etc.) and opportunity, concepts which do not entail

anything about the motivation necessary for the exercise of them.

It is when "a can lift this stone at t" is taken as meaning roughly "a has the ability and opportunity to lift the stone at t"⁴ that it can plausibly be taken as being analysable as "If a wants (tries, chooses etc.) to lift the stone at t, he will lift the stone at t". It is as an analysis of "can" in this "all-in" sense, as J.L. Austin called ([16], p. 229) it, that I shall now attempt to assess the adequacy of the suggested conditional. (Henceforth, I will let "ability" and synonyms cover opportunity as well.)

One trouble with this hypothetical analysis is that it does not make it plain what connection there is between the "can" of ability and the other "can"s I have discussed; it does not explain why we use the same word in both these sets of cases. To establish this link, consider "If a wants to bring about p at t, it is ontically possible that a brings about p at t" - which is equivalent to " $\underline{M}(\underline{s} \ \& \ \underline{a} \ \text{wants to bring about } \underline{p} \ \text{at } \underline{t}) \ \& \ \exists(\text{a proposition } \underline{r}) (\underline{M}(\underline{s} \ \& \ \underline{r}) \ \& \ \neg \underline{M}(\underline{s} \ \& \ \underline{r} \ \& \ \underline{a} \ \text{brings about } \underline{p})) \ \& \ \neg \exists(\text{an expansion } \underline{s}^*) (\underline{M}(\underline{s}^* \ \& \ \underline{a} \ \text{wants } \underline{p}) \ \& \ \neg \underline{M}(\underline{s}^* \ \& \ \underline{a} \ \text{wants } \underline{p} \ \& \ \underline{a} \ \text{brings about } \underline{p}))$ ". Now, "If a wants to bring about p at t, he does" clearly entails "If a wants to bring about p at t, it is ontically possible that he does". Furthermore, in a fully deterministic world, the converse implication also holds, for if all true propositions - taken in conjunction with the proposition that a wants to bring about p at t - are compatible with the proposition that a does bring about p at t, then nothing would explain why a does not bring about p at t, if it were to occur; but in a deterministic world there is, per definition, an explanation of everything that happens. Hence, in a deterministic world "If a wants to bring about p at t, he does so" is materially equivalent to "If a wants to bring about p at t, it is ontically possible that he does so".

My proposal is now that, on grounds of systematic coherence, we take "ontically possible that one does, if one wants" rather than "does, if one wants" as (a first approximation to) the correct analysis on "can" in the all-in sense.⁵ This proposal has, as hinted at, the merit of presenting this use of "can" as an intelligible extension of the "can" of ontic possibility: in fact, the "can" of ability will be a conditional variant of ontic possibility, where the protasis concerns motivational states of the agent. (I would prefer to leave it an open question what the content of the antecedent is when this sort of "can" is predicated of inanimate things.) But this account of the matter also gives a satisfying explanation of why it has seemed plausible to so many philosophers that "If a wants to bring about p at t, he does" is the right analysans: the explanation lies in the natural inclination of human beings to suppose that everything that happens has a

cause, an assumption which makes this analysis materially equivalent to the correct one. It appears to be not an insignificant fact that the traditional form of the conditional analysis has been put forward by philosophers who have assumed the truth of determinism and have sought to harmonize it with the view that there are genuine alternatives of action.

"If a wants to bring about p at t, he can bring about p at t" (i.e. "If a wants ..., it is ontically possible that he brings p about at t") is a genuine conditional, since facts about a's motivational state are undoubtedly relevant when it comes to determining the ontic possibility of his executing some action. That "can, if one wants" is a genuine conditional has usually been denied since Austin made the claims ([16], pp. 209-10) that it seems to resist contraposition and to entail its own consequent. However, both of Austin's claims are false. The conditional does not entail its consequent; it does not entail "It is ontically possible that a brings about p at t" simpliciter. True, there is an interpretation of "a can bring about p at t" according to which the sentence is derivable from the conditional, to wit, when the sentence functions as a recognized ellipsis for the conditional or, in other words, as an assertion of ability (and not of unconditional ontic possibility). By the same token, contraposition does not yield unacceptable results: there is nothing objectionable in "If a cannot bring about p at t, he does not want to". One might feel inclined to object to it on the ground that it "pre-supposes that the absence of a wish is a kind of factor that determines ability" ([18], p. 147). But this objection is based on the false assumption that "cannot" in the protasis means "is unable" (cf. [211], p. 33); the latter can, however, be substituted for the former only when the former is expandible into "cannot even if one wants to", i.e. it is equivalent to this whole conditional phrase. It is in fact not at all absurd to say that one cannot simply because one does not want to, e.g. "a cannot lift the stone now because there is a viper under it (and he does not want to be bitten)". So, the reasons for denying that "If a wants to bring p at t, he can" is a genuine conditional, evaporate when the equivocation of "can" between (unconditional) ontic possibility and ability (=ontic possibility conditional upon a motivational state) is firmly held in view.

My way of modifying the hypothetical analysis dissolves some traditional puzzles. After having suggested that "I could" is short for "I could, if I had chosen" G.E. Moore went on "or (to avoid a possible complication) perhaps we had better say 'that I should, if I had chosen'" ([143], p. 131). Here Moore appears to equate "I could, if I had chosen" and "I should, if I had chosen", taking them both as equivalents to "I could". This seems to have the

absurd consequence, as Keith Lehrer has in fact argued [116], that (to switch to the present tense) "I can" is equivalent to "I shall". Davidson has, however, shown ([68], p. 142) that the conclusion Lehrer draws does not really follow: "I can if I choose" is, Davidson argues, on the hypothetical analysis equivalent to "I shall if I choose if I choose", and the latter is doubtless equivalent to "I shall if I choose", without it following that "I can" is equivalent to "I shall". However, Davidson's argument is not entirely satisfactory in that it goes against the grain to think that the frequently used construction "I can if I choose" means something as artificial as Davidson says it does. I suggest instead that "can" is sometimes an ellipsis for "can if one chooses" and that the latter is merely (given that the world is deterministic) materially equivalent to "shall if one chooses".

2. Since the analysis of the "can" of ability I have proposed retains the hypothetical character of the traditional analysis, it might be held that it inherits some of the difficulties allegedly to be found in the latter. Chisholm has raised (e.g. in [48] and [50]) the following objection (but similar complaints have been voiced much earlier in the history of the subject). Suppose that it is true, not only that the condition that a wants to bring about p at t is a sufficient condition (in the circumstances) for a's bringing it about that p at t, but that it is also necessary:

(1) a brings about that p at t only if a wants to bring about p at t.

There are some types of action - e.g. balancing a knife on the tip on one's nose, pronouncing a tongue-twister - which a human being can hardly ever do except intentionally. But even for the execution of more humble kinds of action, it may be true that an attitude of wanting is necessary - as well as sufficient - on a particular occasion; in fact, it may plausibly be held that unless this is so, the action cannot be executed at will. This is enough for (1). The second step of the argument consists in the assertion that (1) is consistent with:

(2) a cannot want to bring about p at t.

(1) and (2) obviously entail:

(3) a cannot bring about p at t.

But then (1) cannot contain:

(4) a can bring about p at t,

which would be the case if the hypothetical analysis in its traditional form were correct.

Some philosophers - e.g. Aune [12] and Davidson ([68], pp. 143-4) - have responded to this argument by claiming that in (4) a sense of "can", which

is applicable specifically to (intentional) actions, is employed. Now (3) is the negation of (4) only if "cannot" carries a corresponding sense in (3). But this cannot be, these authors argue, since the argument requires that "cannot" has the same meaning in (3) as in (2) - otherwise the validity of the inference is jeopardized - and "to want" does not designate any action.

This reaction is, however, unsound. It begs the question at issue to assume that there is a special sense of "can" applicable to action, for it seems that the only reason supporting this assumption is the view that the traditional form of the hypothetical analysis succeeds in capturing one sense of "can" - and the correctness of this analysis is precisely what is at stake. I have maintained that we can do without postulating any such sense, that the "can" of ability is a conditional variant of the "can" of ontic possibility.

For Chisholm's argument to have any bearing on my position, (1) must be replaced by:

(1*) a can bring about p at t only if he wants to bring about p at t.

If "can(not)" in (1*) and (2) carries the force of ontic (im)possibility, (3) undoubtedly follows, interpreted as asserting that it is ontically impossible that a brings p at t. It seems then that (4) could only be a misleading ellipsis for a judgement that is really conditional, viz. (1*); it appears indefensible as a categorical judgement of possibility.

I think my analysis of conditionals in the preceding section makes it plain what strategy I will adopt in response to this revised argument. According to my theory of conditionals, it is inconsistent simultaneously to assert (1*) and (2), for a construction like "if a wants to bring about p at t" pre-supposes that it is epistemically possible that a wants to bring about p at t. But if one is in a position to assert (2), i.e. to assert that it is ontically impossible that a wants to bring about p at t, then one has thereby shown that the excluding truths are accessible to one, i.e. that it is epistemically impossible that a wants to bring about p at t. That the truth of (2) is not accessible to the speaker - not necessarily that (2) is not true - is a necessary condition for (1*) being true. Contrast this with Keith Lehrer's verdict on the matter:

"The implication /that nothing prevents the agent from trying or choosing to execute the action/ is not, however, a logical consequence of the truth conditions of the conditional statements. It is, instead, a consequence of the conditional statement in conjunction with the usual background assumption that there is nothing to keep a person from trying or choosing to perform an action." ([118], p. 252)

This divergence of opinion between Lehrer and myself regarding the nature of the implication enables me to reject the counter-examples Lehrer presents

to a conditional analysis: e.g. (p. 250) that the conditional "If the agent tries to swallow, he will (can)" is true when it is clear that the agent cannot swallow because he is irrationally convinced "beyond hope and doubt" that he will not succeed in his attempt (and trying logically requires some hope of success).

Given my interpretation, (4) is not - despite its form, which is that of an unconditional judgement of possibility - misleading. For it entails a categorical judgement about mixed possibilities, namely a judgement to the effect that an epistemic possibility that a wants to bring about p at t, and an ontic possibility that he will then succeed in fulfilling his want, obtain.

This may provoke the challenge that epistemic possibility is too brittle a foundation for the doctrine one might call "alternativism", i.e. the thesis that at many times in the life of an agent he can do some other action than the one he actually performs. Does my position not force me to reckon with the possibility that some day the laws and facts of human motivation will become known to the extent that, for every agent on every occasion of intentional action, it will be possible to exclude every attitude of wanting but one? If so, the alternativism I have offered breaks down, since we will never be in a position to assert conditionals of the form "If I want to bring about p then I will (can)" if we know for sure that we will (not) want to bring about p.

However, I believe that cogent reasons have been given for doubting that it is possible even in principle for an agent to formulate a reliable prediction of his own future attitudes and behaviour (cf. [160], [172], chap. 6, sec. 7). I have only space for a sketch of the argument in its barest outline. (5) It is logically impossible that a prediction contains the statement that the prediction itself has been advanced.

This thesis is a special case of the rule that it is logically impossible that there should exist a proposition p such that if a at t thinks (asserts etc.) that p what he has thought (asserted) entails that he at t has thought (asserted) that p. I here take (5) to be sufficiently obvious not to need any further support.

(6) It is always a subjective epistemic possibility (for the predictor), prior to his gaining knowledge of the content of a prediction, that the fact that he is making the prediction he is making will influence his future attitudes and behaviour.

Since, as a matter of empirical fact, a creature, capable of thinking, is also in its attitudes and behaviour influenced by its thoughts, it can never be excluded a priori that further thinking will carry influence. So, if a predictor

in working out a prediction cannot take account of the particular content of his prediction, it follows that he is in no case capable of excluding the possibility that his formulating the very prediction he formulates will make a difference to his future attitudes and behaviour, i.e. (6). From (5) and (6) one can infer:

- (7) It is not logically possible to produce a prediction known (to the predictor) to be accurate - that is, a prediction that takes into consideration all the facts whose relevance to the outcome is a subjective epistemic possibility (for the predictor) - concerning one's own future attitudes and behaviour.

Now to the extent that one's own behaviour affects the beliefs of other people, and thereby their attitudes and behaviour - not to speak of more direct influence on their bodies - this logical limitation of predictability extends beyond the private case.

These considerations put a priori obstacles in the way of an advance beyond epistemic possibility in the question of establishing future attitudes of wanting. Therefore, they ensure the perpetual usefulness of constructions like "If I/you/he want(s) to bring about p, I/you/he will (can)", which, according to the above account, is an expansion of one reading of "I/you/he can bring about p". Some will undoubtedly insist that the latter sentence is true in a stronger sense than this - perhaps that human beings are free, have alternatives of action in the sense of ontic possibility. But I cannot discern any feature of human experience that underwrites this contention. Without doubt, we have the impression that we are free to make up our minds about what to do or want, but this is allowed for by the conception I have offered.

Others may feel inclined to argue that I have placed too severe restrictions on predictability. But it should be observed that I have not contended that it is impossible to put forward predictions that are in fact accurate, predictions that take account of every fact that in fact is relevant, only that one cannot be sure of this beforehand. Nothing prevents one from advancing a conditional prediction like "If p and nothing interferes with this state as described, q"; and if the prediction turns out to have no impact on the situation, this might well be an accurate prediction of the world's actual development.

3. In [18] (pp. 123, 128-35) M.R. Ayers objects to conditional analyses on the ground that whereas "a can bring about p at t" is deducible from "a brings about p at t", something of the form "If q, a (can) bring(s) about p at t" cannot be; hence, "a can bring about p at t" does not entail "If q, a (can)

bring(s) about p at t". There are two possible lines of defence against this argument. First, it might be claimed that, appearances notwithstanding, something of the form "If q, a brings about p at t" is indeed inferable, for the consequent may be so formulated that it entails the antecedent. That this is not a far-fetched possibility in the case at hand can be seen if a distinction between two concepts of ability is drawn, namely, between (a) what one is able to do at will, and (b) what one is able to do either at will or on reflex ... This distinction is important already for the reason that, as a matter of empirical fact, these abilities differ in scope: e.g. a person may be able to shed tears when sad, but not at will; he may be able to straighten a horse-shoe with his bare hands when in a rage, but not if he simply wants to. Now, if one is to define the (b)-type of ability, it appears one has to resort to something like "There is some proposition concerning change, q, having a as subject, such that if q, p is caused at t (and, consequently, it is true that a causes p at t)". And this is indeed derivable from "a brings about p at t", given the analysis of action propounded in chap. 2. Thus, Ayers' argument breaks down if it is interpreted as being directed against conditional analyses of the (b)-type ability.⁶

It is however true that a conditional that stands any chance of explicating the content of a's ability to bring about p at will at t, cannot be inferable from "a brings about p at t". But now - and this is the second line of defence - the trouble for Ayers is that it is evident that the same holds for "a can bring about p at will at t"; at the very least, this obviously requires that we further establish that a wants to bring about p at t and that this want somehow contributes to his action. Therefore, Ayers' objection is no better as directed against the conditional interpretation of the (a)-type of ability. Since the reason for the break-down this time is the falsity of the other premise of the argument, it may be suggested that Ayers' thinking the argument valid is due in part to his not explicitly distinguishing between (a) and (b)-kinds of abilities, something which allows his concept of ability to hover between them.

The second line of defence raises the question whether the addition of the mentioned facts about a's attitude of wanting to the fact of the performance - in other words, whether "If a wants to bring about p at t, he does (can) bring about p at t" - offers a logically sufficient ground for "a can bring about p at will at t". It seems beyond reasonable doubt that the answer has to be negative for a large group of actions. If e.g. a golfer wants to hole-in-one on a particular occasion and, by a stroke of miraculous luck, succeeds in so doing, it does not follow that he was able, even on the occasion in question,

to hole-in-one at will. A distinction is commonly drawn between something one brings off through skill and something one brings off through sheer luck (cf. [172], p. 45). One may be led to obliterate this distinction by the observation that it cannot be drawn for another group of actions. For instance, if the conditional "If a wants to straighten a horse-shoe with his bare hands at t, he will succeed" is true, we may legitimately conclude that a can straighten this horse-shoe with his bare hands at will at t; the suggestion "Perhaps he just succeeded through luck" makes no clear sense here. The same goes for bodily actions such as waving one's arms, jumping, running. The important point is, however, that it has been demonstrated that my analysis of "can ... at will", as hitherto developed, is not rich enough to capture the meaning of the phrase when it is applied to certain sorts of action. The question must now be posed: what has to be added to e.g. "If a wants to hole-in-one at t, he will (can) succeed" to obtain a logically sufficient ground for saying "a can hole-in-one at will at t" or, in other words, to bar the possibility that fortuitous happenings contribute to the success?

It is tempting to take one's clue from the way we go about excluding the possibility that something was a fluke: we ask the agent to repeat the performance, and if he succeeds in complying with this demand (a sufficient number of times), we conclude that he indeed has the skill or knack, but if he instead (repeatedly) fails, we conclude that he was just lucky in the original situation. These conditions of verification constitute, it might be thought (cf. [149], pp. 176-8), the meaning of the "can" of skill.

But I cannot discover any plausible way of building these conditions of confirmation into the meaning of the phrase that is of concern here, namely "is able on a particular occasion". If the suggestion is that it must be generally true that whenever a wants to bring about p in external circumstances exactly like those in which he (perhaps luckily) succeeded, then nothing substantial seems to have been added. If the suggestion is modified to cover also situations roughly like that particular occasion, then it must be asked whether the differences that are permitted are causally relevant for the performance or not. In the latter case the second suggestion is reduced to the first; in the former case it is exposed to the objection that it is conceivable that an agent can do something in one type of situation without having the same capacity even in closely similar situations. Furthermore, it is quite imaginable - though highly improbable - that an agent succeeds in repeating several times something he achieves by good fortune, just by being extraordinarily lucky and without possessing the requisite skill (cf. [33], p. 136 and [154], pp. 189-90). Our reliance in practice on the test of

repeatability as a guide to skill is justified by the fact that with each successful repetition it becomes increasingly improbable that the agent is just lucky, since it becomes progressively more and more improbable that the more or less imperceptible and unforeseeable conditions on which luck depends remain constant or are replaced by other equally favourable ones; but it seems unwarranted to assume that repeatability logically excludes luck. My conclusion is that the proposed analysis confuses meaning and method of confirmation (cf. [33] & [118], p. 251).

If this account is rejected, what is to be put in its place? One rather obvious difference between the agent who succeeds through skill and the one who succeeds through luck is that whilst the former will not be surprised at learning of his success, the latter will very likely be so. The man who possesses a skill - and, hence, knows how to do something - must know what the outcome will be once he has observed that he has succeeded in realizing the initial step. For instance, a golfer's claim to know how to hole-in-one can be justified only if he knows whether he will hole-in-one, once he has observed what his swing was like; he need not attend to the path of the ball. (But it seems to me that it is not necessary that he knows even before carrying out the initial phase that he will succeed if he tries, for knowledge-how does not seem to entail being able: a totally paralyzed person might know how to turn off the light without being able to do so.) I suggest, then, that "a knows how to bring about p at t" is equivalent to "a knows that a conditional of the form 'If one (not necessarily oneself) brings about q at t, then p is brought about by one at t' is true". Possessing a skill could be viewed as knowledge-how in conjunction with ability to make true the antecedent of the conditional.

It is not a necessary requirement that the agent is in possession of the explanation why this conditional relationship holds, i.e. that he is able to state laws and facts in conjunction with which the antecedent entails the consequent; it suffices that he has a soundly based conviction that there are such laws and facts. Note that the conviction must be founded on sound evidence - else the conviction that the conditional is true would not qualify as knowledge. Besides, we can imagine a lucky golfer who is confident about the truth of some recipe (as we might call a conditional of the form in question), without having any good reasons for this confidence. Now even if this recipe happens to be true, it does not follow that the golfer knows how to hole-in-one. The "sound evidence" for the truth of a recipe usually consist in the agent having had experience of repeated attempts and their outcomes, his having noticed correlations between certain initial phases and certain

later happenings.

Is there a perfect match between the applicability of the notion of knowledge how to act and the intelligibility of the question "Luck or skill?", so that we never find one without the other? It may be argued that if I am correct in my opinion that the latter question makes no sense in the case of bodily actions, the applicability of knowledge-how is the wider one. For as we have seen a bodily action is not ontologically basic - neuron-firings, muscle-contractions etc. precede e.g. a finger-movement. So, it might be suggested (cf. [104], pp. 72-4), if one learns this piece of neuro-physiology, one might be said to know how, by what means, one may perform the bodily action of moving a finger. However, as we shall see in 6.3.2, it is part of the concept of means (and, by stipulation, of the antecedent of conditional action-recipes) that the realization of means is epistemically more accessible - i.e. easier to verify - than the realization of what they are means for. But it is evident that so long as perception and proprioception are functioning normally, the verification of macroscopic bodily changes is more easily attained than that of microscopic bodily changes. Since, normally, the verification of bodily changes is immediate, the agent cannot achieve them by applying means, by putting knowledge-how into practice.⁷

Still, the essential point for my purposes is that talk of knowing how to act applies wherever the question "Skill or luck?" can be raised, for this is all that is needed for my analysis. Suppose, then, that the former cannot apply to bodily actions because it asserts the availability of some means, a technique or method; then if it can be shown that such actions can be luckily executed, my analysis of the richer, luck-excluding "can ... at will" has been disproved. But the results of bodily actions cannot be achieved with the help of fortuitous happenings, since this implies "via another route than the one the agent had in mind". And the immediate verifiability of bodily changes is at odds with the agent viewing anything he does as an antecedently verifiable stage on the way to bodily change.

So far as I can see, there is only one way in which luck can creep into the execution of such a relatively basic action: the agent happens to try at a time when he is in particularly good shape, when the physiological pre-requisites in his body are especially favourable, and succeeds thanks to this. But then the agent has the ability in question, at least at the relevant time; the favourable physiological state is precisely something that "ability" is used to designate.

Thus, I conclude that there are two readings of "can ... at will": (i) a thinner one, sufficient in the case of relatively basic actions, such as

bodily ones, for which the hypothetical interpretation offered earlier is adequate, and (ii) a richer, luck-excluding one, required in the case of relatively non-basic actions such as manipulative ones, into which the concept of knowledge how to act - i.e. knowledge of the truth of a conditional action-recipe - enters, along with ability (in the thin sense) to actualize the protasis of the recipe.⁸

Let me stress that in my elucidation of the notion of knowing how to act I have flouted Ryle's dictum ([172], p. 32) that "knowing how" cannot be analysed in terms of "knowing that", for I define "a knows how to bring about p" as "a knows that a conditional of the form 'If one brings about q, then p is brought about' is true". I have no qualms about contradicting Ryle because I think it has been convincingly shown (e.g. [207], pp. 94-105) that the grammatical construction of "knowing how" is dependent on that of "knowing that".

I would also like to point out that this last sense of "can" is connected with the others in that in most circumstances it is ontically impossible that someone lacking in knowledge-how should succeed in fulfilling his want to perform some relatively non-basic action.

4. A summary of the senses of "can" I have been at pains to separate might be useful.

- (1) The "can" of logical possibility.
- (2) The "can"(s) of epistemic possibilities (where "can" (in the subjective case) is a colloquialism for "may").
- (3) The "can" of (categorical) ontic possibility.
- (4) The "can" of hypothetical ontic possibility.

Under this head one can distinguish several species depending on what is the content of the protasis; two particularly important ones are:

- (4a) a protasis encompassing every fact of change pertaining to the agent that is causally efficacious;
- (4b) a protasis mentioning only conditions that make the action voluntary (voluntary ability).
- (5) The "can" of knowledge-how or, more precisely, of knowledge-how in conjunction with (4b) (but henceforth I shall use "knowledge-how" to convey this conjunction).

This list is in no way meant to be exhaustive; I have merely tried to structure the material so far as it is clarifying for my purposes. Nor is the word "sense" to be taken seriously. What I have attempted is to exhibit some distinct zones of a continuum of usages of "can". The continuum is

constituted by the function of expressing compatibility or non-exclusion (cf. [18], p. 176); the differentiating factor is the comprehensiveness of the material of which the consistency is asserted, material which ranges from what is explicitly stated - in (1) - to every true proposition in (3). My grasp of the concept of meaning is not sufficiently firm to allow me to individuate senses with any confidence in this context.

5. REASONS AND WANTING

5.1 The Relation Between Reasons and Wanting.

I have claimed that reasons for action can be put in a conditional form in which, as I have chosen to formulate it (the rationale for this choice is given in 6.3.2), the antecedent is a description of an action-possibility whose realization the reason is a reason for, and where the consequent is a proposition that counts for or against this realization. Thus, a phenomenal reason of a's for or against bringing about p is typically a thought of his expressible by the sentence: "If (or only if) I bring it about that p, q is brought about (by me)". The question I now want to discuss is this: why do some values for "q" turn this into a reason for realizing p, while others transform it into a reason against realizing p, and still others make it neither a reason for nor against? My reply is that the source of this difference is in a's wants vis-à-vis diverse values of "q".

To expound this thesis as clearly as possible, we must first see to it that we are alive to the complexities of negation in regard to a construction like "a wants to bring about q". First, there is the distinction, familiar since chap. 2, between "(wants) to bring about -q" and "(wants) not to bring about q". But we must also keep these, and especially the latter, apart from "has no want to bring about q". When I use formulae like "a does not want to bring about q", they are to be understood in the second way, i.e. the negation pertains to "bring about" not "want". I let "a has no want to bring about q" report both the situation of genuine indifference where a has considered q and -q and found them equally (un)attractive and the situation where a has never thought of q or -q, but I do not use it in the situation where a has a non-ambivalent aversion to q.

Now I can state my thesis on the relation between wanting and having a phenomenal reason for action:¹ a's thought expressible by "If (or only if) I bring about p, q is brought about as well by me" constitutes a phenomenal reason for him to bring about p iff he wants to bring about q; the same thought-content constitutes a phenomenal reason against bringing about p iff he has an aversion against bringing about q, i.e. he does not want to bring about q; and it constitutes neither a reason for nor against iff a has no attitude of wanting towards q, i.e. wants neither to realize, nor not to realize, q. Thus, whether a conditional proposition that forms a content of a's thought is a phenomenal reason for him, and in that case whether it is a reason for or against realizing the action-possibility mentioned in the protasis, depends on a's attitude of wanting towards what is described in

the apodosis.

The following reflection supports this thesis. It seems to me to be a necessary truth that if a conditional is a phenomenal reason for a to cause p, then the fact that this conditional is a content of a's thought must make him more inclined to cause p than he would otherwise be. Surely, the point of practical reasoning is to obtain sufficient motivation for acting, for it is to lead up to (intentional) action. But then those propositions that are relevant in someone's practical reasoning - i.e. are reasons for acting for someone - must make him either more or less disposed to execute a certain action. Now I will argue, in 5.3 and 5.4, that a want is best understood as a certain disposition or tendency to act. If I can show further - as I will attempt to do in 6.1 - how a want can be transmitted from the motivating part of the reason (the consequent) to the motivated one, I will have established the connection between the thesis of the conceptual relation between reasons and wanting and the thesis of the necessary animating power of reasons (i.e. their power to make one more disposed to do the type of action they are reasons for).

But let us assume that these two theses are false, that is, that what is a reason for a person is logically independent of what he wants and, consequently, that reasons do not necessarily animate, and trace out the consequences of this assumption. Kurt Baier [24] has propounded a theory of this type. According to him, the view that reasons necessarily have "motive power" is an illusion generated by a fixation with explanatory reasons (pp. 149, 153-4). Baier's main reason for rejecting this motivational view seems to be that he regards it as incapable of dealing with the fact that people sometimes act in defiance of the beliefs which constitute their best (phenomenal) reasons (p. 142). (I have only room for a brief discussion of this difficulty, see (3) in "Conclusions and Limitations".) Put in a nutshell, Baier's theory of reasons is that "reasons ... are particular facts to which, in particular contexts, universally true (or false) consideration-making beliefs apply" (p. 98). Hence, on Baier's account, consideration-making beliefs - or, as they are also called "rules of reason" - seem to play a role similar to wants in mine. What, then, is a consideration-making belief? Such beliefs (or, more exactly, the contents of such beliefs) are universal "propositions to the effect that if a line of action is of a certain sort then the agent has a reason for or against entering on it" (p. 94). Belief in such a proposition must be widespread in a community for it to generate reasons (presumably for members of the community); rules of reason are "group convictions" (p. 95). Baier enumerates several kinds of rules of reason: individual rules of reason, which prescribe that one has reason to do that which satisfies one's own desires ("self-regarding" rules

of reason, pp. 108-18) or the desires of others ("other-regarding" rules of reason, pp. 118-21); social rules of reason, which proclaim that one has reason to do what is customary (chap. 5); and moral rules of reason (chap. 6), like e.g. that stealing is wrong and that one has reason to avoid it (p. 170). There are also rules of reason that regulate the order of strength among reasons, e.g. "Reasons of law are superior to reasons of self-interest" (p. 100).

Baier appears to be aware of the fact that unless we want to do actions of the kind rules of reason prescribe, i.e. actions of the kind we have reason to do, these rules of reason will have no effect on our actions (pp. 142-3). Why is it that we have such wants? Baier's answer is that it is due to training or upbringing (pp. 147-8). Now this reply cannot mean that our upbringing has bestowed us with certain reasons for wanting to do and actually doing what rules of reason demand, that these wants rest on phenomenal reasons, for that would force Baier to stipulate rules of reason of a second order in order to account for the reasons we have to follow reasons of the first order. And if these second-order rules of reason are conceived on the model of first-order reasons, the same question must rise again, and Baier would be caught in an infinite regress. So it must rather be the case that we necessarily want to do the sort of action rules of reason prescribe for their own sakes (granted that we want to do them at all).

Furthermore, not only must it be the case that these wants are not based on (phenomenal) reasons, it must be logically impossible to justify them by citing (real) reasons. For such a justification would again necessitate the postulation of second-order etc. rules of reason. (Of course, the rules of reason Baier assumes might not be independent, perhaps some, e.g. social and moral ones, are justifiable in terms of others, e.g. individual; but on pain of circularity all rules cannot be so justifiable.) So, if Baier is right on the nature of reasons, we land ourselves in a position where somebody's wanting to do what he has (some type of) reason to do is a contingent fact which is logically beyond justification in terms of reasons, i.e. it is a non-rational want. This certainly seems to be an unpalatable conclusion.

It is interesting to note that Baier appears to think so himself: chapter 12 of his book is dedicated to the task of answering the question "Why should people do what is in accordance with and refrain from doing what is contrary to reason?" (p. 89) - a question which I can only interpret as being equivalent to "What real reasons are there for people to do ...".² It would take me too far to follow Baier's argument and document my reading of it, but my impression is that what is really going on can be summed up in two points: (i) justifica-

tion of all other rules of reason in terms of self-regarding ones (despite this being incompatible with the claim for "the supremacy of moral reasons", pp. 308ff) and (ii) a circular justification of self-regarding rules in terms of themselves (those who abandon these reasons "will not do as well as those who adhere to /them/" (p. 304, my italics), where doing well is presumably conceived in terms of satisfaction of desire). It is not surprising that Baier adopts strategy (i), for of all his rules of reason self-regarding ones are the only ones about which it scarcely makes sense to ask "Why should one follow them?", i.e. "Why should one do that which satisfies one's desires?". This is apparently due to their not being independent of the agent's desires. So, Baier himself feels strongly that on his theory the question "Why should people follow rules of reason?" must be given an answer, and yet he is only capable of providing such an answer because one of his types of reason - self-regarding ones - does not really fulfil the condition of being desire-independent.³

To sum up the argument, the view that reasons are independent of wants creates a dilemma. Because of this independence a question like "Why should people want to do and actually do what they have reason for doing?" can hardly be rejected as illegitimate or meaningless. But "why should people" is naturally understood as roughly equivalent to "what (real) reasons are there for people to", and returning an answer to this question would involve a proponent of the theory under consideration in the regress of postulating an infinite hierarchy of reasons. On the other hand, if he chooses to refuse to allow this question, or the mentioned interpretation of it, then, besides being burdened with the implausibility of this move in itself, he has to concede that, though people's wanting to act and actually acting according to reasons is a contingent matter of fact, it is something that by logical necessity lies beyond what is justifiable in terms of reasons.

A decided advantage of my theory is that it blocks this "why should"-question without having the consequence that people's wanting to act and acting in accordance with reasons acquires an air of non-rationality or caprice. For this is no longer an empirical matter of fact, but a conceptual truth: per definition, what people have (phenomenal) reasons to do is what they want to do (or, if the reasons are real, what they would want to do had they been aware of them), i.e. what they are disposed to do.

Richard Norman has backed up the claim that the content of reasons is determined by public norms rather than by wants with the assertion that even the latter are dependent on norms: "public norms of rational action are prior to the notion of 'wanting'" ([146], p. 73). He draws a parallel to the

phenomenon of language:

"it is as impossible for a man's wants and actions to have a private meaning which makes them intelligible to him alone, as it is for a man to employ a private language intelligible to him alone" (p. 55).

A want gets whatever intelligible content it has from "its relation to public, supra-individual standards and norms" (*ibid.*). These norms and the concepts figuring in them, not wants, fix the content of reasons (e.g. p. 63).

My reply to this is that Norman puts the cart before the horse in taking the intelligibility of a want as dependent on the existence of norms. His view faces the obvious objection that intelligible wants can be ascribed to animals and infants despite the fact that they lack awareness of socio-linguistic standards. Of course, Norman can write off these wants as wants in a secondary sense, as a result of the fact that the behaviour of these creatures can be seen as analogous to the behaviour of norm-conscious human beings (cf. pp. 74, 76-7). But Norman also claims that there is a set of shared "natural reactions" which forms "a precondition of the possibility of rational behaviour" (p. 73) and so presumably of the public norms which define rationality. Precisely in what terms are these natural reactions to be characterized? In the language of pure physical movement? That seems grotesquely inadequate for the foundation of a normative order (cf. 3.4.2). It looks as though nothing less than characterization of the natural reactions in the "intentional" vocabulary of "wanting" will do the trick. But then Norman's position is rendered circular.

After this defence of the doctrine that a person's reasons are dependent on his attitudes of wanting, let me emphasize the precise form this dependency takes. As I conceive it, propositions concerning a's wants - e.g. that a wants to bring about p - are not a part of his reasons. The whole content of the reason is the hypothetical proposition. The relation to wanting comes in when one characterizes a hypothetical proposition as a reason for acting. The reference to wanting is not part of that which is characterized as a reason, but part of what lies in this characterization itself. This account of the content of reasons and of the relation between wanting and having a reason seems to be in substantial agreement with the view of Phillips Griffiths,⁴ whereas for instance Davidson⁵ and Goldman ([88], chap. 3, sec. 6) take the line that reference to the agent's desires is part of the reason (as well as the line that it is the beliefs themselves, rather than merely their propositional content, that is a constituent of reasons).

This difference of opinion may, in some measure, be a reflection of diverging interpretations of the concept of wanting. Treating statements about the agent's wants as a component of his reasons is less tempting when one takes a

dispositional view of wants. But if one conceives - as Goldman ([88], chap. 4, secs. 2 & 3) and Davidson ([71], pp. 43-4, 53-9) appear to - having an occurrent desire as having a thought of a certain content, a parallel treatment of wants and beliefs in reasons appears inevitable. And if one further concentrates on the use of "reasons for action" in contexts of explanation (rather than evaluation), it is realistic to assume that it is the facts that the agent has certain wants and beliefs, i.e. thinks thoughts of certain contents, rather than these thought-contents themselves, that constitute the reasons, for it is the former that explain an action.

An exclusive focus on explanatory contexts may also encourage this form of reasoning: a reason-explanation of an action is not complete unless wants are mentioned alongside beliefs; therefore, reference to wants must be part of the reason itself.⁶ But this is a non sequitur analogous to: a causal explanation is not complete unless a causal law is cited together with the specification of the initial conditions; therefore, the law is itself a causal condition. It is quite gratuitous to assume that everything set out in a reason-explanation forms a part of the reason (when "reason" is not a mere synonym to "explanation"), just as it is to assume that everything cited in a causal explanation is a causal condition.

As I see it, the phrase "is a reason for" designates, both in the practical and in the theoretical sphere, a relation of support. In the practical case, this relation of support consists in an attitude of wanting directed towards what is described in the apodosis of a hypothetical: it is wanting that turns a thought-content of conditional form into support for an exemplification of a particular action-type the possibility of which is presupposed in the protasis. In the theoretical dimension, the support consists just in the conditional relationship: that p is a reason for thinking that q just means that a conditional like "If p, (probably) q" is really true (real reason) or believed to be true (phenomenal reason). Thus, when a thinks that q because he thinks that if p then q and that p, his phenomenal reason for thinking that q is only his thinking that p. The reference to his thought of conditional content comes in when one characterizes his thought that p as his phenomenal reason for thinking that q. The presence of this hypothetical thought (or belief) is what distinguishes reasoning from p to q from merely associating q to p and locates the former process on the level of reflective thought.

It follows from the dependency of reasons for action on attitudes of wanting that whatever is a reason for causing p is a reason for wanting to cause p.⁷ This is also what a dispositional analysis of wanting would have one believe, since it seems clear that whatever is a reason causing p must be

a reason for being disposed, or tending, to cause p. The conversion, that whatever is a reason for wanting to cause p is a reason for causing p, looks equally plausible. For any reason for wanting p may be (one of) the reason(s) for which one actually brings about p, given that contrary reasons are sufficiently weak or collaborating reasons sufficiently strong. But since the phrase "one's reasons for causing p" does not imply that one has actually caused p, there seems to be no reason for withholding it unless an action is produced; rather, the thing to say is that the reasons for wanting p are, independently of the production of p, reasons for causing p. So we reach this principle:

(P) Whatever is a reason for causing p is a reason for wanting to cause p and vice versa.

I affirmed above that a phenomenal reason for bringing it about that p is typically expressible in the form "If (or only if) I bring about p, q is brought about (by me)", i.e. that p is conceived as a sufficient (or sufficient-and-necessary) condition for something (which is wanted). But obviously, that p is thought to be only necessary for q - i.e. a conditional of the form "If I do not bring about p, q is not brought about (by me)" - can also give one a reason for causing p. Therefore, some weight must be placed on the inserted qualification "typically". What is intended by it? A conditional of the latter sort can only enter into a's practical reasoning if it is an epistemic possibility for him that there exist states of affairs such that if (and perhaps only if) he brings them about in conjunction with p, q is brought about as well. In other words, practical reasoning in accordance with a necessary-means-conditional can be construed as reasoning in accordance with a gappy sufficient(-and-necessary)-means-conditional "If (and only if) I bring about p and ..., q is brought about (by me)", where completion is an epistemic possibility for the reasoner. The full explanation of why the reasoner must think it possible to fill in such a conditional comes in section 5.3, when it is established that wanting something logically presupposes thinking there might be a sufficient condition for achieving it.

A spectacular difference between practical and theoretical reasoning is this. If a wants q and thinks that if he brings about p then q, he has some reason - though a reason that can be outweighed - for bringing about p (contrast [219], p. 31); but if he thinks that q and that if p then q, then he has no appreciable reason for thinking that p (unless he in addition thinks that the number of sufficient conditions for q is very limited). Another way of expressing the same point is this, that in practical reasoning, the relevant attitude - namely, wanting - is to some extent transferable from con-

sequent to ground, whereas the relevant attitude in the theoretical dimension - viz. thinking - is not so transferable. This is one important datum to be explained by an analysis of wanting.

The results of my protracted investigation into the concept of a phenomenal reason for action can be codified in a definition like this:

(PR) "... is a phenomenal reason for a to bring about p" is true iff: (1) the blank is filled by a proposition expressible by "If (or only if) I (i.e. a) bring about p (perhaps along with ...), q is brought about (by me)", (2) this conditional proposition is the content of a thought of a's (he need only think that the conditional is probably or possibly true), and (3) a wants to bring about q.⁸

In order to obtain a definition of a real reason, (2) must obviously be exchanged for a clause demanding that the conditional be true. Moreover, (3) ought to be somewhat weakened by tacking onto it a clause of this sort "or a would want to bring about q if he thought of q and (some of) its consequences". The second adjustment is obviously connected to the first.

5.2 The Object of Wanting.

1. I have tacitly assumed above that the objects of wanting can be specified by means of a "that"-clause prefaced by "to bring about" (etc.). I would now like to explain and defend this assumption.

Wanting is wanting-to-do something, i.e. (according to the analysis of action in chap. 2) wanting-to-cause something to become a fact or wanting-to-bring-about the instantiation of some property in a spatio-temporal region. The hyphens are meant to indicate that "to cause" (etc.) is not part of the object of wanting, but an expansion rendering explicit something inherent in the notion of wanting. This assertion about wanting can be taken in at least two different ways. First, as a statement that whenever the verb "to want" is used in an utterance there is a logically equivalent paraphrase that employs "to want-to-cause" instead. Secondly, as a statement that this is sometimes so, and a stipulation that I shall here concentrate exclusively on those cases where such an expansion is permissible. I intend the assertion to be taken in the latter way. (The reason why I fasten on the indicated subset of uses of "want" is that they alone appear to be pertinent to the analysis of reasons and reason-governed actions.) In non-technical discourse, "to want" seems sometimes indistinguishable from "to wish", and in these cases the former is scarcely expandible into "to want-to-cause". (The distinction between wanting and wishing will be discussed in 5.3.1.) From a systematic point of view, it would be desirable that people abstained from employing

"to want" when "to wish" would do equally well - thus perhaps making the original assertion true in the first interpretation. But as things stand, expansion into "to want-to-cause" can only serve as a test that we have a use of "to want" that is of present interest.

Let me illustrate in more detail how this test works. The verb "to want" can be followed by three types of object-construction:

- (i) a noun: "to want an apple";
- (ii) a verb-construction in the infinitive form; here we must make a subdivision between (a) the case where the verb is an action-verb ("to want to move a finger", "to want to climb a mountain") and (b) the case where it is not ("to want to die", "to want to forget something humiliating"); and
- (iii) a propositional clause: "to want that the bully leaves one alone" (or more naturally: "to want the bully to leave one alone").

It seems reasonable to hold that case (i) can be reduced to case (iib), that e.g. "to want an apple" is equivalent to "to want to get (have) an apple". Now neither in the case of (iib) nor in the case of (iii), does there appear to be any difficulty in eliciting the state of affairs the wanting must be a wanting-to-cause (to obtain) for the wanting to be relevant to my concerns: wanting to die is simply wanting-to-bring-about that one dies, wanting the bully to leave one alone is wanting-to-bring-about that the bully leaves one alone, etc. If these transcriptions are felt to be inappropriate, this is probably because "wanting" is read as passive, powerless wishing. But if anyone should doubt whether "to want" is ever expandible into "to want-to-cause", I cannot do better than appeal to the theory of wanting, to be developed in the following two sections, according to which a want is a disposition or tendency to cause.

In the case of (iia), however, we have an option of two ways of rendering what is wanted-to-be-caused: "to want to move a finger" e.g. can be construed either as "to want-to-cause that a finger moves" or as "to want-to-cause that one moves a finger". On the first alternative, what is to be caused consists in the fact that a certain bodily movement occurs, on the latter in that an action of the wanter occurs. I opt for the first manner of conceiving the wanted object: when in a construction of the type "to want to F" "F" takes an action-verb (according to the definition in chap. 2) as value (or a compound beginning with such a verb), the state of affairs forming the object of wanting is the intrinsic result of the action, not the state of affairs consisting in that the wanter commits an action having this result. I shall soon try to justify this standpoint, but first I have to take care of what

appears to be a more radical objection against the manner in which I construe the object of wanting, to wit, as being expressible in a "that"-clause. This will provide an opportunity for making the purport of this view more precise.

Jerome A. Shaffer asserts "that sexual desire is not a case of desiring that" ([189], p. 178), that to sexually desire an object is not to desire e.g. that one have sex with it. However, this claim is ambiguous. It could be a denial of the claim (i) that the objects of wanting or desire can always be rendered by "that"-clauses, or of the claim (ii) that they can always be rendered by intensional "that"-clauses. (i) follows directly from my explication of the wanting of relevance here as wanting-to-cause and my view of the relation of the causal relation (see 2.3). On this interpretation, the context "wanting-to-cause that ..." will be just as extensional as "the fact that ...": different descriptions of identical particulars and universals can be substituted salva veritate. Now it is obvious that what Shaffer has in mind is not the rejection of (i) but of (ii), for he goes on to maintain that sexual desire "does not consist in the envisaging of some state of affairs toward which one has a positive attitude" (*ibid.*, my italics). Furthermore, he complains that the theory he criticizes cannot accommodate the extensional features of sexual desire (pp. 181-2).

If this reading of Shaffer is correct, I need not take issue with him: (ii) is indeed false. Sometimes a desire is simply a tendency to reflex-behaviour: sexual desire is a case in point, for sometimes the behaviour which is the manifestation of such a desire is clearly reflex-behaviour: the desires to withdraw from a source of pain and to eat when hungry are other examples. In these cases, the behaviour appears to be sparked off along fixed lines by awareness of certain bodily states (tension in the genital areas, damage of tissue, the contractions of an empty stomach): there is no need for the agent to envisage what is to be done. The context "want-to-cause that ...", if understood this way, is not intensional, just as the phrase "causes that ... on reflex", which describes the manifestation of the desire, is not intensional. I am, however, concerned with wants or desires which are conceptually linked to reasons and which can culminate in intentional behaviour. Since the context "intentionally causes that ..." is intensional - more precisely, is governed by the constraints that flow from (T6) in 3.2.1 - "wants to cause that ...", when describing the underlying desire, is so as well.

Let me not create the misapprehension that whether the context "wants to cause that ..." is intensional or extensional is necessarily revealed by superficial inspection of what fills the blank. The desire to cause that one's limb is withdrawn from a source of pain may be a tendency to reflex-

behaviour, may be automatically triggered off by a sensation of pain, but one may also desire this for the reason that it will put an end to the pain, i.e. desire it in an "intensional sense". The same may clearly be true of sexual desires and the desires connected with hunger and thirst. Whether the context "desires to cause that ..." is intensional or not, and the ensuing behaviour intentional or not, depends on whether the agent's thought of the object is involved or not. The force of this remark is far from clear, but more or less the rest of this work is dedicated to spelling it out.

To sum up. I endorse (i), but I do not embrace (ii). It is only true of a subclass of desires that the "that"-clause expressing their objects is intensional; it is only true of those desires that, given suitable circumstances, manifest themselves in intentional behaviour. It is these desires that are focussed upon in this work.

2. I now turn to the question concerning what type of proposition or state of affairs forms the object of wants (of the intensional variety) in cases of sort (iia). Contrary to me, Kenny has affirmed this general rule: "Desires of the 'wanting to ϕ ' form may be regarded as ... wanting oneself to ϕ , wanting that oneself should ϕ " ([111], p. 124). That is, Kenny seems to take the view that, in the case " ϕ " takes an action-verb as value, the object is constituted by an action-proposition, i.e. a proposition consisting in that the wanter performs this or that action, rather than, as I hold, propositions consisting in that the intrinsic result of these actions obtain. Thus, according to Kenny, wanting to move a finger is roughly wanting that one moves a finger, whereas, in my opinion, it is wanting that the finger moves.

Now if it is granted that the "wanting" we are concerned with here is expandible into "wanting-to-cause", Kenny's position looks untenable, for e.g. wanting to move a finger is obviously different from wanting-to-cause that one moves a finger. A characteristic way of satisfying the latter want is by doing something - e.g. putting the finger on something hot - that causes one to move it, but this is a most uncharacteristic way of satisfying the former want. But perhaps Kenny would not accept this unfolding of "wanting" into "wanting-to-cause"; so let us cast about for an argument that does not turn on this assumption.

An action-description can be modified by adverbs or prepositional phrases like "intentionally", "involuntarily", "at will". So, if a's wanting to move a finger presently is his wanting that he moves a finger presently, it should be possible to prod him to make the content of his want more precise by asking e.g. "Do you want that you presently move your finger at will or not?". Note

that I am not saying that a must either want that he moves the finger intentionally or want that he moves it non-intentionally before the question is asked, since an actual action is either one or the other. Such a reasoning would not be correct, because indeterminacy is a recognized feature of intentional objects (i.e. what occurs in the intensional contexts of mental verbs). Rather, I am arguing that he is faced with the option of making the object determinate in certain respects. But suppose one modifier, e.g. "at will", were impossible in the context of "to want ..."; then the complementary one could be said to be implicit, i.e. a's wanting that he moves his finger would be indistinguishable from his wanting that he moves his finger non-intentionally. Now it would really be paradoxical if a's wanting to move a finger presently turned out to be indistinguishable from his wanting that he non-intentionally moves a finger presently, for suppose a finger-movement would result from this want in a way that made it intentional, then a's want to move his finger would not be satisfied!

We must then ask: is it logically possible for someone to want that he presently performs an action at will? This is, of course, only possible if it is possible for him to have the second-order want that he presently has a want to perform the action, since the performance of the action will be done at will only if it results in a certain way from a want to perform the action (as we shall see in chap. 6). Is it then possible to want that one presently has the want that p?⁹ It seems to me very doubtful whether this is possible, because I think the following two statements are evidently true:

- (1) Whatever is a reason for wanting that one wants p is a reason for wanting p; and
- (2) There can be no genuine want satisfaction of which is logically guaranteed whenever it is present.

(1) is a corollary of (P) embraced in 5.1,¹⁰ while (2) will have to be accepted on its own credentials. On the basis of (1) and (2) it appears one can construct an argument for the conclusion that there can be no second-order wants of the kind wanting that one presently wants p. For if the reasons a at t is in possession of give rise, at t, to a want that he at t wants p, and (1) holds, it seems that we should expect a want that p to be produced as well at t. (The only thing I can think of that would block this is if a, though capable of forming the second-order desire, were incapable of forming the first-order desire. But the only meaning I could give to "incapable" here is "lacking the requisite conceptual equipment", and it is obvious that whoever has the equipment for forming the second-order desire is also equipped for forming the first-order desire.) So, if a at t were to

have a want that he at t has a want that p, he would inevitably at t have a want that p. In other words, his second-order want would be automatically gratified whenever it arose - but that violates (2).

The upshot is that it is impossible to want that one presently has a want, hence impossible to want that one presently performs an action at will. But then if wanting presently to F - even when "F" represents a verb of action - is wanting that oneself F-s presently, it is wanting that oneself F-s non-intentionally presently. That however, as we noted above, is absurd. So, wanting to execute an action at present is not wanting that oneself performs this action presently. Since the only other serious candidate for the role of the propositional content of wanting is states of affairs consisting in the intrinsic results of action, I conclude, substantially in agreement with Prichard,¹¹ that wanting to F, where "F" represents verbs of action, is wanting(-to-cause) the result of F-ing. As Annette Baier puts it, speaking of intentions: "the agent's special privilege is, not to insert himself into all his intentions, but to leave himself out of the basic ones",¹² i.e. the agent does not figure as agent in the content of intending (wanting).

Two cautions should be inserted here. First, although I have claimed that it is logically impossible to want that one presently has a want that p, nothing I have affirmed precludes the possibility of someone wanting that he presently wants p partly for some reason R*, distinct from the one he already has for wanting p, R. Since, ex hypothesi, R is distinct from R*, the second-order want is not here automatically gratified. I think that Harry G. Frankfurt's case ([82], p. 9; cf. also [173]) of a physician who wants to have a desire for some narcotic in order sympathetically to understand his drug-addicted patients can be construed as of this sort. The physician does not merely want to have a desire for drugs, he wants to desire them for the same reasons as his patients, e.g. to relieve a state of bodily tension. Because this second-order desire involves the physician in wanting to be in this state of bodily tension, there can be no risk of the satisfaction of his second-order want being logically guaranteed.

Note, however, that this concession does not impugn my argument against action-propositions being the objects of wanting. The concession that it is possible to want that, one presently wants to act for some reason one does not already have beforehand, does indeed suggest that it should be possible to want that one performs an action at will (partly) for this reason in the present. But it could scarcely be held that "to do F at will for some specific reason R*" merely makes the content of the want to perform F more precise: it evidently introduces further changes to be caused, i.e. those that make R*

a reason for one to perform the action (changes which consist either in creating an attitude of wanting towards the apodosis of R^* or making it a credible hypothetical).

The second caution is that it is feasible to want that one wants p at a future time sufficiently distant for it to be possible that one's present attitudes of wanting have altered. (Henceforth, this time will be referred to as "future".) If a wants at t_1 to have a want at t_2 that p , then, according to my reasoning above, he must at t_1 want that p , but if the distance between t_1 and t_2 is sufficiently long for a want to change between them, it does not follow that a at t_2 as well has the want that p . So, a 's second-order want is not here trivially satisfied. As a consequence, the argument I marshalled above against interpreting e.g. "wanting to move a finger presently" as "wanting that one presently moves a finger" does not work against understanding "wanting (presently) to move a finger in the future" as "wanting (presently) that one moves a finger in the future". But surely a uniform interpretation of "wanting to F presently" and "wanting to F in the future" is to be desired. Therefore, I propose that the latter equally means wanting-to-cause the result of the action, F -ing, e.g. that wanting (presently) to move a finger in the future" is equivalent to "wanting-to-cause (presently) that a finger moves in the future".

As a matter of fact, this wanting-to-cause that a finger moves in the future can take two forms. In the first, one restricts oneself to wanting-to-cause presently what are prerequisites for one's performing the action of moving the finger in the future, i.e. what is necessary for one's having, in the future, the ability to move the finger, should one so desire (then). Here there is no want to influence one's future motivational states, one's attitudes of wanting: one relies on their remaining constant unless information one considers pertinent as well at the present time turns up. In the second case, one wants to influence one's future wants as well, i.e. one strives presently to set up a sufficient condition for one's moving the finger in the future. As a more realistic illustration of these two forms, contemplate somebody who wants today to catch an early morning train tomorrow. He may rest content with wanting to buy tickets today, to wind the alarm-clock today etc., i.e. with wanting to do things which are necessary and sufficient for his being able to catch the train tomorrow. But he may also want to affect his wants of tomorrow - e.g. by rehearsing his reasons for catching the train so that he will be capable of recalling them in his drowsy state in the early morning - if he suspects that, when he wakes up, still sleepy, in the dark morning, his want to catch the train might otherwise be overpowered by a

desire to stay in bed.

3. After having discussed second-order desires, a brief comment on second-order reasons may not be out of place. My main objective will be to sketch an account of second-order reasons that fits the general account of reasons given in 5.1 and to argue against the suggestion that second-order reasons are of a different breed from the one the general account circumscribes.

Higher-order reasons, as I conceive them, are simply reasons for or against deliberating about some action, viz. reasons for or against thinking about reasons for and against some action, or reasons for or against suppressing, not taking account of in acting, the outcome of some piece of deliberation already engaged in. Since deliberation can be undertaken as a (voluntary) act (although it need not), it is not to be wondered at that there can be action-reasons for and against engaging in it. As I shall indicate below, these reasons have the general form expounded in 5.1.

Joseph Raz has propounded ([169], pp. 35ff) a theory of second-order reasons that diverges from mine. He seems to take second-order reasons to be reasons not concerning the act of deliberating about an action, but concerning the action that is the object of deliberation. For him, second-order reasons are reasons which concern the same action which the first-order reasons concern, but which overrule the first-order reasons whatever their strength. In one of his examples (pp. 37-8), he considers someone who is invited to consider a business-proposal - which is open just this one night - but declines to do so for the reason that "she is too tired and upset to trust her own judgement" (p. 37). Raz seems to be claiming that we must regard this reason "as a reason ... for rejecting a business proposition despite the fact that her weariness does not bear at all on the merits of the proposition" (p. 38). If this is correct, there must be reasons which cannot plausibly be expressed in my cherished conditional form.

But I do not think it correct. The weariness is a reason for forbearing from engaging in the action of calculating the merits and demerits of the proposition - a reason easily expressible in my standard-formula: "If I consider the proposition in this state of weariness, it is very probable that I would make some grave miscalculation". It is true that, as the example is set up, abstaining from carrying out this calculation is tantamount to making it impossible for one (rationally) to accept the offer. But it does not follow from this that the reason for abstaining from the calculation is a reason for rejecting the offer. First, because, although on a Davidsonian view of action-individuation (see [66]) the action of refraining from consideration of

reasons is identical to the action of making it impossible for one (rationally) to accept the offer, the latter must not be confused with the action of rejecting the offer. Secondly, even if the last-mentioned identification were allowed, substitution of co-referential phrases in the context "is a reason for ..." is not permissible. On the strength of these two grounds, I conclude that what is undoubtedly a reason for refraining from calculation is not also a reason for rejecting the proposal.

5.3 The Concept of Wanting: Basic Features.

1. As I have intimated, I will defend a dispositional account of the concept of wanting. Let me therefore start by stating what I take to be the major consideration in favour of this type of theory. It is a well-known fact that states of wanting can be of greater or lesser intensities. Now, what is the criterion for one want being, say, stronger than another? There could, it seems, be no criterion other than that the former desire is victorious - i.e. manifests itself in behaviour - in a situation where (only) these desires compete. According to this view, it is a conceptual truth that in a class of conflicting desires the strongest one is the one that expresses itself in behaviour. If this is so, the strength of a state of wanting is presumably the strength of a disposition or a tendency to act. The rival account that holds that there is merely a contingent connection between the intensity of wanting and the probability of behaviour-manifestation encounters the formidable difficulty of specifying what the strength of a want is the strength of, if it is not the strength of a disposition to act. One traditional type of account - attributable perhaps to Hume¹³ -, that the intensity is the intensity of an internal impression or feeling, will not do for the simple reason that, however cleverly elaborated, it will be phenomenologically untenable: it is obviously false that we experience some feeling whenever we want something. (I will later on say something more positive about one way in which feelings may accompany desire.)

I take it, then, that the prima facie credibility of the dispositional type of analysis is very great, and I therefore state as a thesis about wanting:

(W1) "a wants p" ascribes a disposition or tendency to act in some manner to a.

The trouble starts with the task of specifying exactly what sort of disposition is being ascribed. How, for instance, is one to distinguish between this disposition or tendency and the disposition or tendency to vomit that a sick man is afflicted by? They must clearly be distinguished, for although there

may be some sense in which it is true that the sick man wants or desires to vomit, I am interested in the more restricted sense of "wanting" in which it is appealed to in the explanation of acting at will, and vomiting is as a rule not something voluntarily caused.

Further conceptual requirements of wanting must then be sought. In [111] Kenny advances the claim that the object of wanting must be logically possible. His reasons are (p. 121) that wanting "finds expression in two ways: verbally, and in behaviour", but that both these forms of manifestation are excluded in the case of a desire with a logically impossible object, since "what is logically impossible is indescribable" and "there are no steps towards a logically impossible end". Kenny anticipates an objection to the effect that it is enough if the agent thinks that there are steps to his end. He attempts to counter this by affirming that it is not adequate "for whenever we try to specify the object of his want, we are condemned to utter nonsense or to contradict ourselves" (p. 122). I find it hard to understand what could be meant by the last statement and the earlier one about the indescribability of what is logically impossible. Surely, if I attribute a contradictory opinion to someone, I must be prepared to answer his legitimate question about which contradiction he is committed to by describing it in words. This does not force me to contradict myself, since "a holds that p" is not automatically turned into a contradiction if p is contradictory (not if p is an implicit contradiction). Moreover, it is a patent fact that people frequently want what is in fact logically absurd, e.g. mathematicians have wanted to square the circle. Kenny says that these people are mistaken concerning the objects of their wants, but if they do not want what they say, then what could they plausibly be portrayed as wanting?

Since wanting is, after all, a psychological attitude, it seems far more reasonable to assume that the restrictions on the content of the attitude must be of a psychological origin, e.g. that a person can (logically) want only what he thinks is logically possible. I am, however, ready to go further than this and assert that the wanter must think it within the scope of his ability - understood in the manner of (4) in 4.2.4 - to bring the content into existence at the relevant time. More precisely, the thesis I want to defend is this:

- (W2) It is logically necessary that, if a, at t, wants-to-cause p at t then he must, at t, think that a conditional of the following type might at least be true (i.e. the truth of the conditional must be an epistemic possibility for him): if he is properly motivated to cause p at t (i.e. there is a change such that if he is subject to it at t), it is ontically

possible that p be caused at t by him.

I have given a's thought that he has the ability in question the weakest possible form: that of subjective epistemic possibility. A realistic account cannot, I think, demand more, for it appears possible that someone may want to realize something he thinks he has only the slightest chance of accomplishing.

In the light of (W2) the interlocking of the two components of a reason - a conditional thought and an attitude of wanting directed towards its apodosis - begins to be discernible. If a thinks that, if (or only if) he brings about p then q, it is guaranteed that the necessary condition laid down by (W2) for wanting q is satisfied. For, according to the analysis of conditionals in 4.1, this thought of a's presupposes that he regards it as at least epistemically possible that he brings about p - which splits up into it being an epistemic possibility that he has the ability and the motivation to bring about p. But if (W2)'s restriction on wanting p is fulfilled, its restriction on wanting q must also be fulfilled, since causing p is thought sufficient for causing q. Furthermore, the way for a transferral, in practical reasoning, of an attitude of wanting from apodosis to protasis, is prepared: we need not fear that such a transferral will violate (W2)'s constraint on wanting directed toward the antecedent.

One feature of (W2) is worthy of special notice: the ability referred to in the thought is not voluntary ability, but ability of the more inclusive type, i.e. not (4b) but (4a) of 4.2.4. Talk of what a thinks he can do at will, i.e. if he wants to, would generate a circle and would thus make (W2) useless for the purpose of elucidating the notion of wanting. (The fact that the reference to wanting in the definiens would occur in the intensional context of "a thinks that ..." does nothing to rescue (W2), for one must make room for the possibility that a wanter gains insight into the notion of wanting, e.g. comes to realize that wanting something involves thinking one might have the ability to get it, without being caught in an infinite regress.) This is not meant to suggest that a must think that his ability is of a non-voluntary kind; rather, the point is, that to the extent that the thought that one has a certain capacity is presupposed by one's having an attitude of wanting, this thought must be indeterminate concerning the aspect of voluntariness/non-voluntariness. It is a commonplace that thoughts exhibit the feature of intensional indeterminacy: in thinking e.g. that a certain man wears a pair of shoes, one need think neither that they are size 10 nor that they are not; one's thought may be indeterminate concerning the size of the shoes.

Austin once pointed out¹⁴ that the logical truth of the conditional "If

a wants to bring about p, and a can bring about p, he brings about p" to an equal degree warrants the conclusion that "a can bring about p" is equivalent to "if a wants to bring about p, he brings about p" and the conclusion that "a wants to bring about p" is equivalent to "if a can bring about p, he brings about p" - which is why it in reality warrants neither. For someone espousing a dispositional account of wanting, there is a perpetual temptation to subscribe to both these analyses, thus moving in a full circle. Stuart Hampshire may be an example of an author who has given in to this temptation. On the one hand, he says quite explicitly: "'A wants to do X' is indeed equivalent to 'other things being equal, he would do X, if he could'" ([99], p. 36). On the other hand, several passages seem to document his acceptance of the traditional conditional analysis of "can": e.g. that (in a context where the "will to succeed" is not assumed) inquiry into your power is inquiry into "what you would succeed in doing if you really wanted to and really tried" (p. 30). (Cf. also what he has to say about the testing of powers, e.g. pp. 23, 29.) This appears to be what he sums up in the statement that "'want' and 'can' ... depend on each other for their sense" (p. 17). Now, it may be that two concepts have to be directly defined in terms of each other, but we should not resign ourselves to this as inevitable without having done our utmost to find explanations that make at least one of the concepts independent of the other. So far my own account avoids, as I have endeavoured to show, this circularity; but I have still not come round to spelling out what kind of behaviour the want to cause p is a disposition to, and it might well be thought that reference to voluntary ability will be necessary at this point, since it may seem that the behaviour can hardly be anything but that of causing p, which will take place only if the voluntary ability to cause p is present.

Before enlarging upon this, I must, however, put forward a couple of considerations in support of (W2). First, the plausibility of (W2) depends, in a large measure, on the notions of wanting and wishing being carefully separated. The presently relevant point of difference is that whereas wanting is directed towards states of affairs which the wanter thinks that he might be able to realize, wishes take as their objects propositions the realizations of which the wisher definitely places outside the range of his power.¹⁵ For this reason, it is a common phenomenon to find people wishing they had acted differently in the past, but no one wanting this. However, on my account there is nothing logically absurd about the idea of a person wanting something in the past to be different - not, that is, if it is granted that there is no explicit contradiction in the assumption that it is

in an agent's power to affect the past (and the fact that some philosophers believe in the intelligibility of the notion of backwards causation seems to prove that it is not an explicit contradiction). For what an agent can logically want to do is, in my view, a function of what he can logically believe himself capable of accomplishing. So the rarity - not to say total absence - of utterances like "I want to smoke a cigarette yesterday" is, on my account, due only to the pervasiveness of the conviction that what is past is past and cannot be interfered with in any way.

But if one holds that the past cannot be interfered with, one is easily persuaded that one cannot interfere with the present either - if the reference of "the present" is understood very narrowly; one's power of influence is restricted to the future, immediate or more distant (cf. [142], pp. 20-1). Therefore, if precision is desired "a wants at t to bring about p at t" - where "t" must be taken as referring to a time having duration - should be exchanged for "a wants (beginning) at t to bring about p in t-" , where "t" denotes an instant¹⁶ and "t-" a span of time stretching from t to some suitable later instant. (What is suitable varies from action to action, depending on how rapidly the action can be performed.) So, "a wants at t to cause p in t-" will be my way of rendering the idea that a wants to cause p straightaway, i.e. (roughly) at the time the want obtains. To achieve this equivalence, "t-" , when it appears in the context of a propositional attitude, must be understood to stand in for a description such that the agent cannot fail to realize that it refers to his present, e.g. "now", or more strictly, "the very next moments".

Someone will probably be inclined to object against these temporal constraints that (a) we commonly say such things as "I want to do what I am presently doing" and that (b) only someone excessively ascetic could require that people should stop wanting something the moment their want is gratified, thus denying them the joy and pleasure ensuing from the fulfilment of a want.¹⁷ Both objections are, however, equally easy to rebut. Ad (a): I think that "I want to do what I am presently doing" is most naturally construed as synonymous to "I want to go on doing (roughly) the same sort of thing I am presently doing". The correctness of this interpretation is evidenced by the fact that if the addressee interrupted the speaker in his activity, the speaker's response would naturally take some form like "Didn't you hear what I said?", "Why won't you ever respect what I want?" etc. Ad (b): though I have asserted that it is irrational (if logically possible) to want what one is convinced is the case, I have certainly not embraced the view that it is irrational to be glad or pleased about what one is convinced is the case.

The latter follows from the former only if one analyzes having the emotion of joy or pleasure that p as containing having the want that p, and I would definitely reject that. (Something more on the relation between having emotions and wanting will be said in 5.4.5.) I can quite consistently maintain that wanting p transforms into being pleased that p at the moment one becomes aware of having achieved that p.

A second consideration in support of (W2). If a person by making certain exertions becomes fully convinced that he cannot bring about a certain state of affairs, this conviction will cause his attempts to cease; he will not - like insects which lack the intelligence to learn about their incapacities through experience - go on vainly exerting himself until he collapses of fatigue. Now although other explanations are possible - e.g. that the want is suddenly superceded by a conflicting want - the most likely explanation is surely that the behaviour ceases to be wanted. (If, however, the behaviour was wanted as a means to an end, the end might still be wanted, provided that the agent thinks that there might be other means to it.) Hence, wanting p has a precondition in the thought that it might be possible for one to cause the materialization of p.

However, human beings sometimes engage in certain forms of behaviour even though they are aware of their futility: someone being tortured may desperately pull at his ropes to get away from the source of pain, in spite of the fact that he is utterly convinced that the ropes will not give way. In such a case it is natural to speak of a desire to free oneself from the bonds and to remove oneself from the source of pain. But the desire here is a tendency to reflex-behaviour (i.e. is an extensional desire): the pulling at the ropes is, just like screaming and grimacing, a reflex elicited by awareness of a present or future pain-sensation. I am, however, seeking to define the meaning "to want" has when it is used to designate a disposition to act at will. (W2) is, then, only intended to be valid for wanting in this - the intensional - sense.

2. So much for the explanation and underpinning of (W2). The task of finding a set of conditions necessary and sufficient for wanting must now be resumed. Can such a set be constructed out of (W1) and (W2)? If a wants p, then, since a want is a tendency to act, it is natural to think that it is thoughts of p which tend to move a to action, and, more precisely, to actions which consist in making it true that p. The thought that one might be able to bring about p is a thought in whose content the proposition p is embedded; so this thought ought also, on this hypothesis, to have a motivational influence. This suggests the following simple way of tying together the ideas of (W1) and (W2): a's

wanting to bring about p is logically equivalent to his having the thought that he might be able to bring about p, and this thought tending to cause p. However, it is evident that this does not capture a sufficient condition of wanting. Suppose a is nauseated; then his envisaging in some way - perhaps, by accident, as embedded in the indicated modal frame - the proposition that he is vomiting might cause him to vomit. Still, it does not follow that he wants to vomit (in the "intensional sense") and that his vomiting occurs at will.

A diagnosis which lies close at hand is that this proposal is defective because the fact that the proposition p is presented to a under the modality of something he might achieve is not utilized in the explication; the modal frame appears to be an inessential ingredient, a mere ornament. An idea put forward by C.H. Whiteley provides a clue of how this modal aspect of a's thought can be utilized:

"if a man desires to do x, it does not follow that, even in appropriate circumstances, he will do x, but only that he will do what he supposes to be x ..." ([212], p. 48).

Now one can only think that what one is causing is (or might be) p if one thinks that one is (or might be) capable of causing p. So if it is demanded that a's thinking that he might be able to cause p tends to cause, not necessarily what is in fact p, but what a thinks is or might be p, the modal aspect of a's thought is exploited.

But is it merely a coincidence that a's thought that he is able to cause p tends to cause precisely what a thinks is p? No, this should clearly be a matter of causal necessity: a's ability-thought causes whatever a thinks of as p; if a had applied "p" differently, what he would have tended to cause would have been correspondingly different. It is not that what a's thought that he is able to cause p tends to cause and that what a thinks is p vary independently, and that they merely by a happy accident coincide. Incorporating this refinement, we reach this proposal: a wants to bring about p iff a's thinking that he might have the capacity to bring about p tends to cause what he, non-coincidentally, thinks might be that p. In short, a want (in the "intensional sense") is a disposition or tendency to cause something which is non-coincidentally viewed as the result of an action one (more or less tentatively) ascribes to oneself the ability of performing.

Let us see if we are now in possession of a condition strong enough to exclude extensional desires, i.e. tendencies to reflex-behaviour. Consider behaviour occurring on reflex, e.g. as a response to physical pain. First we must confront the question whether this behaviour is directly geared to the sensation or whether recognition of the sensation in thought, some attention

paid to the sensation, is an essential mediating link. It seems to me that, whereas the former is the case on phylo- and ontogenetic levels, so primitive that the notion of identifying sensations in thought is not applicable, the latter is the case on levels where this notion applies, for it seems that reactions to pain disappear or at least subside when one's attention is distracted from the sensation. Let us, in order to get as close a parallel as possible to intensional desires, keep to the sophisticated level. Thus, I assume that pain-reflexes are not (directly) caused by pain-sensations but rather by thoughts of the form "I am now feeling pain".

Now note something characteristic of reflexes: one may become aware of having performed them only post factum, if at all. This is most conspicuous in the case of reflexes which involve a change of a minor part of the body and which are rapidly executed, e.g. blinking - when one blinks on reflex one may remain totally unaware of the fact that one has blinked. Hence, it is clearly false that someone blinking on reflex must think of what he is doing as blinking (or as anything else). This is equally true of grosser pain-reflexes, like e.g. withdrawal of a hand, though due to the grossness of the movement one will normally, by observation, become aware of its taking place sooner or later. Reflexes are a form of automatic behaviour, i.e. behaviour the execution of which is not supervised by thought, behaviour which is not engaged in because it is thought to be of a certain kind. Therefore, it is not necessary that someone who withdraws an arm on reflex thinks that he has the ability to withdraw his arm. Nor does the thought that elicits the behaviour - the thought expressible by "I am now feeling pain" - in any way presuppose that the agent attributes any ability to himself. Of course, the agent may think that he has the ability, but this would be a mere coincidence; the modal aspect of this thought has no role to play in the genesis of the reflex. So, an extensional desire, a tendency to reflex-behaviour, is definitely not a tendency to behave in a way that is non-coincidentally viewed as an exercise of an ability assigned to oneself.¹⁸

This conception of (intensional) wanting deserves, then, to be stated more rigorously:

(W3) a wants at t to bring about p in t- =df

a thinks at t that he might (at least) be able to bring about p in t-, and this thought tends to cause what a, at t, thinks will (possibly, probably) become p in t-, and the ability-thought would not have tended to cause what it in fact tends to cause in t-, if the latter had not been thought of by a as p.

(W3) is composed of the following three ideas: (i) a's thought that he

perhaps has the ability to cause p tends to cause something to become a fact, (ii) a thinks that what is tending to become a fact is p, and (iii) it is no coincidence that a thinks this of what is becoming a fact. On the basis of (W3), it is easily seen why the context "wants to cause that ..." is intensional: this intensionality derives, by virtue of (ii) and (iii), from the intensionality of "thinks that ...". (i) enters into the definiens because it is required by (ii).

The Whiteleyian manner of identifying the behaviour towards which there is a tendency has one further advantage. To speak of a causal tendency is to speak about what would actually be caused in a certain type of hypothetical situation, and this manner of identification enables one in principle to spell out the details of this situation without generating a circle by having recourse to the idea of a voluntary ability, which would have been necessary if wanting that p had involved the tendency to cause what is in fact p. What a thinks is p will actually be caused by his thought that he can cause p if this thought is maintained sufficiently long for the causing to be completed. Several things may prematurely end this state of thinking: (a) a's mind goes blank (he dies, faints etc.), (b) something turns up which causes something a regards as incompatible with p, and (c) a becomes convinced that what he originally thought was p is not p, that he in fact cannot cause p.

The following situation is conceivable: a's thought that he can cause p initiates the causing of what a takes to be p; but then a suddenly thinks that he can instead cause q (which he holds to be incompatible with p), and this initiates the causation of what he takes to be q. In this situation a fact which appears "truncated" as compared to p will have been caused when the thought of q enters a's mind (assuming that what a views as p in fact would have been p). Suppose that "p" = "a's left arm rises to shoulder-level"; then what is actually caused before the interfering factor (a's thought that he can cause q) breaks upon the scene might be slight muscular tension, an almost imperceptible upward motion of the arm. I believe that it is situations of this sort we describe by sentences like "I felt a desire (an impulse) to lift my arm" or "I felt like lifting my arm". It does not seem unlikely that it is sensations of such muscular tensions that the Humean type of theory misconstrues as the desire itself and as a cause of behaviour.

However, though I regard (W3) as casting considerable light on the nature of wanting, it is still far from sophisticated enough to be able to cope with the richness and subtlety of the notion of wanting as we find it in ordinary discourse. In the next section, I turn to some refinements that, I hope, remove this deficiency.

5.4 The Concept of Wanting: Refinements.

1. (W3) creates the impression that wanting that p must be wanting that p for its own sake, because it attributes to the thought that one might have the ability to cause p, to this thought per se, a tendency to cause behaviour. Hence, it is an objection to (W3) that we ordinarily say that someone wants something, even though he does not want this as an end in itself, but only for some reason, e.g. "a wants a box of matches" is correct even when a wants the matches only for the reason that he then will be able to light his cigarette.¹⁹

To accommodate this fact that attitudes of wanting may be based on reasons, (W3) must be replaced by:

(W4) a wants at t to bring about p in t- =df

a thinks at t that he might (at least) be able to bring about p in t-, and this thought - perhaps due to other thoughts of a's of the form "If (only if) I bring about p in t-, q will (probably, possibly) be brought about" - tends to cause what a, at t, non-coincidentally thinks will (probably, possibly) become p in t-.

This should make it clear that the motivating power need not come from the content of p alone, but may also derive from propositions a associates - logically, causally etc. - with p.

The question whether reasons for action are causes of action has been widely debated. By the above addition to (W3), I enlist, but not without reservations, on the causalist-side, with those who argue that reasons are causes.²⁰ (The reservations remain even if reasons are understood as causes, not of actions, but of their intrinsic results.) For it must be remembered that (phenomenal) reasons, as I conceive them, are the content of thought, not facts that thoughts with a certain content occur, and it is quite obvious that it makes no sense to say that a thought-content is a cause of anything. Thus, the non-causalist position embodies a considerable portion of truth. But there are things to be said on the causalist side as well. When causalists have claimed that reasons are causes, they have usually meant (Davidson is a case in point) that thoughts or beliefs are causes. (Their mistake only consists in the view that thoughts or beliefs are constituents of (phenomenal) reasons.) And this causal claim is true.²¹

I am unable, however, to produce a positive argument that proves this contention, but the onus of proof is on the non-causalist: since causation is a well-entrenched concept with numerous indisputable applications, all the causalist is obliged to do, in order for his hypothesis to receive the blessings of simplicity, is to work out his hypothesis in detail and refute

counter-arguments to it - a task to which this thesis is dedicated.

However, the above addendum to (W3) paves the way for further complications; the opposite complaint that the definition in the new shape has become too broad may now be voiced. Suppose e.g. a cashier hands over his cash to a robber only for the reason that, as he believes, the robber will otherwise carry out his threat to kill him; is it then true that the cashier wants to hand over his cash to the robber? (W4) licenses a positive answer, since there is no doubt that the cashier's thought that, if he hands over his cash to the robber then he will save his life, causes him to do what he (correctly) considers to be handing over his cash to the robber. But some philosophers - e.g. Hampshire ([99], p. 40), Danto ([59], pp. 186-7), Aune ([13], pp. 59-60) and von Wright ([219], chap. IV) - have disputed this, claiming that we in ordinary parlance draw a distinction between what we want to do, on the one hand, and what we are forced or obliged to do, on the other. Hence, saying that the cashier wants to give the robber the money implies that he does not really need to be forced, perhaps because he secretly collaborates with the robber.

But other philosophers - like Goldman ([88], pp. 50-4, 105-7) and Audi ([9], pp. 389-91) - have taken the same stand as I. In support of this position, one can point to the Kantian principle that someone who wants an end must, as a matter of logic, want whatever he regards as a necessary means to that end,²² and that the cashier does want to survive and does believe he can do this only by giving in to the gunman's threat. It is apparently true of the clerk that he wants to hand over the money to the robber because he wants to survive and believes that he can do this only by handing over the money to the robber. Moreover, it is not hard to imagine cases where it would be perfectly natural to say of the cashier that he wants to give the money to the robber. Suppose several cashiers were threatened collectively, but were in a position to consult each other and to vote on the matter; they could then undoubtedly formulate the issue in the question "Who wants to give money to the robber?", and our cashier would naturally express his inclination by "I want to give money to the robber".

The upshot is then that we face two conflicting sets of linguistic data, one of which must be explained away somehow. I think it is both systematically more advantageous and intuitively more attractive to accept the latter set at face value and to try to explain away the former. It is systematically more advantageous because it makes the analyses of the concepts of wanting and acting at will simpler. And it is intuitively more plausible too, I think, since what, after all, is forced in the cashier-case, if not the cashier's

will? That is why we would describe the clerk as not acting of his own free will (though he doubtless acts at will). Surely, the difference between this kind of compulsion and purely physical compulsion - as when one moves the limbs of another person's body just as one moves inanimate matter - is that in the former case one influences the motivational state of the agent, his attitudes of wanting.

How then explain away the common contrast between what we want to do, on the one hand, and what we are forced or obliged to do, on the other? Often when constructions of the form "a wants p now" are employed in ordinary discourse, they have roughly the force of "a wants p now either for its own sake or for its consequences in the normal or natural course of events". This interpretation arises, it may be surmised, because, in the interest of greater informativeness, the use of "a wants p now" is restricted to occasions where a's attitude of wanting is based in the indicated way; on the occasions when a wants p because of the non-standard and artificial grounds provided e.g. by obligation or compulsion, it is replaced by "a is forced (obliged) to bring about p now". Violation of this convention results in misleading suggestions, but not downright falsity, since the objectionable character of "a wants p now", in the latter type of situation, can be neutralized or "cancelled" by the addition of "but not for the usual reasons, only because he is under obligation (threatened) to the effect that ...".²³ So, I suggest that the above contrast is in reality between what we want to do for its own sake or for its normal or natural consequences, on the one hand, and what we want to do only because of the peculiar considerations introduced e.g. by a threat or an obligation, on the other.

If this account is on the right track, we are permitted to say of the cashier both that he wants to hand over the cash to the robber (because of the threat) and that he does not want this (because of the normal consequences of giving money to a robber). This depicts the cashier as the victim of conflicting wants - a depiction which is surely correct. But we need a way of marking which attitude of wanting is dominant, which one is formed in the light of all the cashier's beliefs. Let us speak of "wanting_d" ("d" for "dominant", i.e. stronger than every other want of the agent's having an admittedly incompatible object) and "wanting_r" ("r" for "restricted (to normal grounds)"). The cashier then wants_d to give the robber the money, although he wants_r not to, for I have portrayed the cashier's desire to survive as stronger than his normally based aversion to giving money to a gunman. It is wanting_d that is of particular significance for my purposes and that (W4) - so long as no restrictions are placed on the content of the consequents of the reason-conditionals -

seeks to define.

2. Objects of wanting that the wanter takes himself to be able to make true only indirectly, via making other propositions true, raise an obstacle for (W4). There is no problem if a holds a definite view about how to bring about p indirectly, if he believes that he is able to realize p by means of realizing q: a's wanting p is then his thinking that he might be able to bring about p (by causing q), and the tendency of this thought to cause a state of affairs of a kind (to wit, q) that he thinks will make it true that p. But suppose that a has not reached a settled opinion about how to cause p, suppose that he merely thinks that there is or might be some way of achieving p. (This is quite sufficient for it to be logically possible that he wants p.) In this state of ignorance concerning means, it makes no sense to attribute to a a tendency to bring about a state of affairs which he thinks will make it true that p, because there is no specific state of affairs he takes to be of this sort. Hence, (W4) ceases to apply.

I propose that a's wanting p here takes the form of a tendency of a's to preoccupy himself with the question of how to bring about p, which includes, primarily, a tendency to ponder this question, but in some cases also a tendency to reflect on related questions like how to obtain information relevant to answering the first question, and to engage in behaviour necessary to obtain this information. So, the nature of wanting would be seriously misrepresented if we assumed it to be of necessity a "single-track" tendency, manifesting itself only in behaviour thought to be desire-fulfilling. In fact, the manifestations can be multifarious, although the axis is still the manifestation thought to be desire-fulfilling: other manifestations only occur in so far as they are regarded as necessary for this manifestation. To comprise this multifariousness, we must replace (W4) by:

(W5) a wants_d at t to bring about p in t- =df

a thinks at t that he might (at least) be able to bring about p in t-, and this thought - perhaps due to other thoughts of a's of the form "If (only if) I bring about p in t-, q will (probably, possibly) be brought about too" - tends to cause what a, at t, non-coincidentally thinks will (probably, possibly) become p in t-, or the ability-thought tends to cause that a is preoccupied with the question of what means there are of bringing about p.

The quest for means can end in two ways: a finds sufficient means for p or he does not. If the latter, he can no longer want p; if the former, he must want_d some set of sufficient means, otherwise his attitude towards p

will no longer satisfy either limb of (W5).

The fact that someone has a certain want may, then, explain the thoughts he has. Thus, Stephen E. Norris is correct in maintaining that my wanting something could be "a reason why I did any calculating at all about how to obtain it" ([147], p. 83). But "reason why" is here synonymous with "explanation why". Therefore, Norris goes wrong when from the fact that the reason here is not a premise in any reasoning he apparently concludes that reasons for action need not be such premises (pp. 83-4). There cannot be action-reasons for calculation occurring spontaneously, i.e. having thoughts of a certain content cropping up in one's mind, simply because this is no action. Of course, "(mental) calculation" can also designate a voluntary act(ion), namely that of causing thoughts to occur. But Norris has done nothing to show that calculation done at will, for reasons, is not preceded by a piece of reasoning wherein the (phenomenal) reasons appear as premises.

Hitherto I have only considered attitudes of wanting which exist in time-stretches that to a greater or lesser extent overlap or at least border on the times of their object-specifications, namely wants that begin to exist at t, from which the time-span t- stretches. But (W5) provides me with the conceptual resources for dealing with wants the objects of which are to be realized in the more distant future. In such cases a want is a tendency to consider whether there are preparations and preconditions which have to be realized presently and, if there are thought to be such, a tendency to realize them. That is, no considerations new in principle are introduced, as compared with those expressed by (W5), although (W5) does not strictly cover them.

3. According to the account of wanting that has emerged so far, a want can be no causal condition of the result of the relevant action: the causal conditions of the result, to the extent that they are designated by mental terms, lie entirely in the agent's thoughts about what he is able to do and about what the thing he is doing is or will be. Since to ascribe a want to an agent is just to say that certain of his thoughts tend to be causally efficacious, it would, in the words of Don Locke, be

"an error to think that the cause of an action will include wants as well as beliefs, analogous to thinking that any cause must include causation, or a power to bring about effects, as well as the particular factor which is the cause" ([123], p. 176).

So, it appears, a want is no causal condition of the intrinsic result, along with the thoughts, because it is only the tendency of these thoughts to cause the result.

This view is, however, exposed to an objection, which Kenny launches against it: it does not permit "beliefs and wants to be independent explanatory factors" ([112], p. 119). Locke himself comes close to seeing the difficulty. For though he at one point a bit incautiously states "that beliefs, by themselves, can be causally sufficient for action" ([123], p. 176), he acknowledges that there is "the problem of why certain beliefs move some people to action but not others, or move some people in one way and others in another" (*ibid.*) and that

"if an agent's behaviour is to be explicable at all, there will presumably be some factor, independent of the various beliefs which constitute his reasons, whose presence in the one case explains why he acts for those reasons, and whose absence in the other case explains why he does not act, despite having those reasons" ([123], p. 179).

But he expressly denies (p. 176) that invoking wants is a way of solving this problem, a way of indicating the explanatory factor which supplements the beliefs.

If, however, "wanting" does not designate this additional explanatory feature, then hardly any other term of non-technical discourse could be held to do so, and consequently Locke would be driven to the conclusion that it is not possible to frame explanations in non-technical vocabulary which make human behaviour fully intelligible. For such complete explanations we would presumably have to await a further development in neuro-physiology. But this goes counter to what seems to be a common intuition: that explanations in terms of wants and beliefs possess a considerable degree of self-sufficiency, that they do not stand in need of completion by the introduction of e.g. physiological terms (though explanations couched wholly in physiological terms may perhaps replace them). Therefore, it would be a good thing if we were allowed to say that reference to attitudes of wanting is a way of indicating these additional explanatory factors.

The strategy I will adopt to make room for this assertion has been already adumbrated in 3.1.2 where I briefly discussed the nature of dispositions. There I opted for Armstrong's "realist" conception of dispositions, according to which a disposition is an internal state which is identified by reference to the responses certain forms of stimulation evoke. To be concordant with this, (W5) should be modified to:

(W6) a wants_d at t to bring about p in t- =df

a is in an internal state such that his thinking, at t, that he might be able to bring about p in t- - perhaps due to other thoughts of a's of the form "If (only if) I bring about p in t-, q will (probably, possibly) be brought about too" - tends to cause what a, at t, non-

coincidentally thinks will (probably, possibly) become p in t-, or such that the ability-thought tends to cause that a is preoccupied with the question of what means there are of bringing about p.

My use of the terms "disposition" and "tendency" may lead a reader to think that I consider them interchangeable: have I not for instance sometimes called wants "dispositions" and sometimes "tendencies"? True, but this is justified by an ambiguity in "to want" rather than by "disposition" and "tendency" being synonyms in my exposition. To attribute a disposition to something is to attribute an intrinsic or non-relational property to it, but a property that is identified in a relational way, namely by the fact that its instantiation causes a certain effect when influenced in a particular way. An imputation of a tendency goes one step further in that it entails something about the prevailing circumstances as well, to wit, that they are such as to call forth the response if no countervailing factor turns up too. So, an object that is brittle, i.e. disposed to break easily, during its whole existence tends only to break when in suitable external circumstances, i.e. when knocked or struck. Applying this terminology to wants, a want, in the sense defined by (W6) is not only a disposition, but also a tendency, to behaviour, for factors - namely, the ability-thought and perhaps phenomenal reasons - which, unless prevented, will elicit the response, are actually at hand. However, there is also a purely dispositional use of "to want", which I shall not attempt to define, since it is secondary to the tendency-use and of no particular importance in the context of this work. I presume that merely exchanging the dispositional "believes" for the episodic "thinks" in (W6) goes a long way towards giving this definition.

If this account is correct, wanting that p is a state - probably internally describable in neuro-physiological terms - which one might enter into when one begins to think that one has the capability to cause p and which, together with an actual instance of such a thought, tends to cause the materialization of something one thinks is p. Since an instance of this state is clearly an entity distinct from the result of an action that might ensue, there is one good reason the less for refusing to accord it the status of a causal condition for the intrinsic result. True, this state is identified relationally, by means of the effects it tends to have, but this is of no consequence for its status as a causal condition, as long as it is also describable in a way that displays its distinctness from the putative effect.

This position is located in-between two other common standpoints. On the one side, there are those - e.g. Charles Taylor ([198], pp. 33, 38, 49-51), Stoutland [195], Locke [123] and Abelson ([1], pp. 38-42) - with whom I agree

in treating a desire as a disposition or tendency, but who diverge from me in taking this as precluding that desires are causal conditions of actions or their results. This divergence might be due either to their adopting a "phenomenalist" view of dispositions or tendencies or to their misconstruing the requisite independence between cause and effect. On the other side, we find philosophers - like Churchland [55] and Goldman ([88], pp. 72-6) - who believe not merely that wants are causes of actions (I shall not here make any fuss about the point that the effects are specified as actions rather than as intrinsic results), but also that there are causal laws to be discovered which relate beliefs and wants, under these descriptions, to actions.²⁴ This latter view is again an illusion, comparable to belief in the existence of causal laws connecting brittleness, impact and breaking. For to designate a property by a phrase of the type "to want ..." is not to designate it in any way that brings out what is causally relevant about the property i.e. what it is intrinsically like; employing this phrase merely assures us that there is some property that is causally relevant. A state of wanting must be internally redescribed - presumably in physiological terms - in order for the phrase picking it out to be eligible as a constituent of a causal law of behaviour. What makes concepts like wanting and brittleness useful for explanatory purposes in prescientific contexts - namely, their combining the roles of indicating the existence of a causal factor and providing a covering generalization - is precisely what renders them unsuitable for the scheme of scientific explanation, where these roles are split up, in the name of greater informativeness.²⁵

Against this David Pears has urged in [155] that the description by which a desire is designated in a causal law must be available, known, to the agent, and so could not be one only discoverable by further research. His most important reason for this appears to be that an agent's foreknowledge of his intentional action "would be unintelligible if he could not predict his action inductively from his desire identified under an independent, generalisable description" ([155], p. 103). I will in chap. 6 try to undermine this reason by developing an account of an agent's knowledge (correct view) of what he intentionally does, which does not base this knowledge on inferences from awareness of desires. Later in this section, I will also attempt to shed light on knowledge of one's own present desires. Here I shall limit myself to arguing that Pears does not succeed in carrying out his self-imposed task of providing a plausible instance of a description that is logically independent of the effect, generalizable (i.e. fit for entering into a causal law) and available to the agent. In this way I hope to cast doubt on the feasibility

of this task.

Pears' suggestion for such a description of a want_d to do A is this: "the psychological state of the agent which is nearly always the state that produces" ([155], p. 111) "(t)he agent's sincere statement that he all things considered wants to do A" (p. 108). Observe that this is not an internal description of the psychological state, but a functional or relational one. Pears appears to hold - and here I would agree - that an internal description of a desire is not possible at present: a desire "is always judged by its effects, because it is a dispositional entity with no non-dispositional basis that is yet known to us" (p. 132; cf. 115).

A difficulty for Pears is presented by the fact that a sincere report is an episode with a representative content: something is asserted or reported. What? Of course, that one wants_d to do A. But what does that mean? That one is in the state that nearly always produces the sincere report "I want_d to do A"? But how is one able to ascertain that one is in this type of state if one is unable to identify it under some internal or intrinsic description? Surely not by observing that the mentioned effect - one's sincere report - is produced, for that would imply that every sincere report that one presently desires something was preceded by another sincere report to the same effect. Does the report then mean that one is in the state that nearly always (when the ability is present) causes the agent's doing of A, the other type of manifestation Pears ties a desire to (pp. 103ff)? (This is in fact the line Pears took in a personal conversation.) But how does one know that one is in this state, if one is incapable of identifying it under an internal description, the instantiation of which one has learnt is followed by the action? Certainly not by observing its effect on one's action, for Pears supposes that this report can be made before the desire manifests itself in this way: as I have said, he thinks that the agent's foreknowledge of what he will do is based on his awareness that he is in the state that tends to produce the sincere report.

Pears could instead contend - and in fact seemed to do so in conversation - that we cannot rationally justify our avowals or reports of this content, but that there is a non-rational mechanism that allows us to make them: we are simply neurally "wired up" in a fashion which enables us to report what behaviour tends to issue from our current state. But surely the hypothesis that nature is considerate enough to provide us with a non-rational way of reliably reporting something whose truth is beyond rational justification (such as support in what is immediately experienced) is fanciful enough to require stronger backing than anything we have seen so far. It seems then

that, on Pears' premises, there is no convincing way of construing the tendency to make the autobiographical reports in question.

A further source of difficulty is a potential ambiguity in talk of a "generalizable description". Pears seems to require that for a cause to be describable in a generalizable way is for it to "be describable in a way that assigns it to a class each of whose members is followed by a member of a single class to which the effect belongs" (p. 101). But taken in this sense, a generalizable description is a notion which is considerably weaker than that of a description fit to enter into a causal law, i.e. a description which expresses a causally relevant property. For suppose it is true of all times t that iff p obtains at t then this causes q to obtain at t_1 and r to obtain at t_2 , but that the latter effect is not brought about via the former, i.e. we have an instance of branching causal chains. Then, relative to the effect r , the description of p as "the state of affairs whose existence causes q " is generalizable (in Pears' sense) (and logically independent of the effect r). Note that it is also possible to predict r from knowledge that the description "the state of affairs whose existence causes q " applies (given causal laws implying that q and r have a common cause in p). Still, this description of p is inappropriate in a causal law that explains r , for, ex hypothesi, p 's being a state of affairs whose existence causes q is causally irrelevant to its causing r .

Now notice that Pears' reason for demanding that the generalizable description of the desire be known to the agent - that the agent have foreknowledge of his actions - only requires generalizability to be construed in the weaker way, i.e. as not entailing that the description is causally explanatory. If Pears claims only availability of a description generalizable in this sense, his claim does not necessarily conflict with the main line of my argument, since my concern is with causally explanatory ones. However, Pears seems prepared to assert (at least he did so in private communication) that the description he has produced is causally explanatory as well. This claim is fully parallel to the thesis that "the state of affairs whose existence causes q " is a causally explanatory description of p in respect of r , in our just mentioned example: think of " p " as standing for the state of desire identified in intrinsic (probably neural) terms, " q " for the avowal and " r " for the appropriate behaviour. This should make patent the falsity of Pears' claim.

So, it can safely be concluded at least that Pears has not succeeded in isolating a causally explanatory description which is available to the agent, even if it be conceded (as I have argued on p. 118 that it should not be)

that his description is generalizable in the weaker sense of purely regular succession.

4. In the usual course of events, people have knowledge of their present attitudes of wanting. Some philosophers - e.g. Abelson ([1], chap. 2), Alston ([2], pp. 329-34), Audi ([8], pp. 5, 9) and, as we saw in the last subsection, Pears ([152], [155]) - have felt the bond between believing, or being prepared sincerely to state, that one has a certain want, and actually having it to be so strong that they have given the former a criteriological or other role to play in the explication of the notion of wanting. Quite apart from other objections (some of which are touched upon in subsection 5), I cannot see how these views can be justified in view of it being a blatant fact both that people not infrequently deceive themselves about their present wants and that they do not reflect upon their psychological states at all, having neither true nor false beliefs about them. (Both situations can be covered by the rubric "unconscious desires".) So far as I can see, the facts do not permit one to go beyond this statement: as a matter of empirical fact, for most times t , a at t wants p iff a at t is convinced that he then wants p . It ought to be added that this conviction seems to be based on what, broadly speaking, can be called introspection, that is, that a person seems to be able to "read off" his present state of wanting straight away, without any laborious process of inference from a wide range of inductive material.

Every theory of wanting must needs be called upon to explain these epistemological facts. And my theory might be attacked precisely on the ground that it cannot be reconciled with the ease which generally characterizes the way we gain knowledge of our present wants. For my theory requires that a certain causal judgement - to the effect that some thoughts have a causal power - be established in order for a judgement about wants to be, and since the days of Hume it has often been taken for granted that even a singular causal judgement cannot be established by a single observation, but demands a wider inductive backing. So, the challenge (cf. e.g. [88], p. 98) I must try to meet is: if a judgement about what one presently wants entails a singular causal judgement, how is it then possible - as often appears to be the case - that such a want-judgement be advanced, apparently with full justification, solely on the basis of introspection of what happens on a single occasion?

My reply consists in rejecting or, more accurately, restricting the Humean principle of the non-observability of singular instances of the causal relation: I am willing to grant its validity for perception of phenomena outside one's own body, but I find it untenable for proprioception or introspection.

Let me first consider the case of proprioception, lest I be accused of giving in to the temptation of creating special rules for the awareness of mental phenomena. It seems to me that the tactile and proprioceptive sensations I normally have when e.g. an alien object presses against my body and causes it to move conclusively verify the singular causal judgement that the object caused my body to move. (Of course, proprioceptive impressions can, like all sense-impressions, be hallucinatory; but the question at issue is what information about the world they logically guarantee, granted that they are veridical.) There is a marked contrast here to the case where one is seeing, as it is colloquially put, one external object causing another to move: what we see does not exclude the possibility that the second object begins to move independently at the very instant the first reaches it and, hence, that its movement is not at all caused by the movement of the first object. But in the proprioceptive case, there is a world of difference between the sensations I have when an external object causes my body to move and the sensations I have when my body begins to move independently as an alien body comes into contact with its surface; in the first case, I literally feel how the object forces my body to move, how my body gives way to this foreign impact. I find it in no way absurd to say that these experiences, provided that they are veridical, conclusively verify the causal judgement, including its counterfactual implication.²⁶

Not only does it in this manner appear to be possible to verify propositions about what is actually caused; it also appears to be possible to establish that something tended to be caused. If I, when the object begins to exert its pressure, make an effort and succeed in resisting its impact, the sensations I normally have of this seem to verify the statement that the object tended to cause my body to move.²⁷ Given that the empiricist view of concept-formation is roughly true, i.e. that "all concepts are derived from experience", it might be claimed that we in experiences of the sort illustrated possess the empirical substratum from which our concept of a cause is derived. (I would like to stress, pace a writer like von Wright who asserts (see above 2.2 under (5)) that the concept of causation is based on that of (intentional) action, that, as the first example above clearly points out, these experiences are not necessarily of ourselves as sources of action, whether intentional or not.)

Now if the existence of impressions, that verify judgements to the effect that an alien body causes or tends to cause one's own body to move, is conceded, the hypothesis, that propositions to the effect that thoughts cause or tend to cause bodily movement can be directly verified in proprioception and

introspection, should not be rejected at the outset as a desperate and fanciful move. But is this hypothesis true? Phenomenologically, this situation is considerably less transparent, so I think that the strongest support will have to come from circumstantial evidence. Consider a reflex-action like somebody's jumping because of a sudden loud sound in the vicinity. I think few philosophers - even amongst those who are anti-causalists concerning reasons - would dispute these two theses: (i) the jumping movement is caused by awareness of the sound, and (ii) the agent can report immediately that the awareness of the sound was the origin - i.e. the cause - of the movement. These two assumptions strongly support the view that there can be immediate knowledge of instances of psycho-physical causation. Against this background, it is reasonable to interpret the common immediate knowledge of present wants as a special case of this immediate knowledge of psycho-physical causation or causal tendencies.²⁸

However, immediate knowledge of psycho-physical causation does not completely explain knowledge of wants, since an assertion that a want obtains also entails a claim about an additional explanatory factor, a categorical basis of a disposition known only by its effects. This claim - alternatively formulated as the claim that the relevant set of thoughts would not always elicit the same response when they occur in the mind of organisms of comparable capabilities - cannot be verified by observation of one instance. But it is instead a particular instance of a generalization to the effect that thoughts are not by themselves sufficient to move to action - a generalization which has massive support in human experience and which has been built into the meaning of "wanting". The truth of this generalization is, I suggest, something simply taken for granted by a user of the want-terminology on the occasions when he uses the terminology.

Note that there is no inconsistency between maintaining (a) that a person, as a rule, can tell straight off what he presently wants, though this entails a causal assertion, and emphasizing, as I did in the foregoing subsection, (b) that extensive research may be necessary for establishing the description under which a desire enters into a causal law. For, as Davidson has pointed out (e.g. [62], p. 91-2), knowing that a singular causal judgement is true entails only knowing that there is a law which covers the instance at hand, not acquaintance with the concepts of the law.

An account of how people can read off their present states of wanting constantly runs the risk of stepping into the opposite trap of making unconscious desires impossible. I think my explanation avoids this error as well, for there is no necessity for people directing their attention to what they

are concurrently thinking, which of course is essential for determining which thought causes what. Moreover, if a large number of thoughts are entertained just before acting, it may be difficult to separate the causally efficacious thoughts from the causally idle. Thus, there are probably avenues of error in determining present wants that are closed in the case of determining present beliefs.

5. I have proceeded on the assumption that it is possible to set forth a set of logically necessary and sufficient conditions for the truth of sentences of the form "a wants p". Let me now put this assumption under fire by scrutinizing an approach that denies it. I have in mind a kind of approach which parallels one we came across earlier when studying the concept of belief, namely one which has it that the concept of wanting is a theoretical concept, whose content derives from the concept's function in some laws or lawlike propositions that constitute the core of a common-sense psychological theory that underlies ordinary usage. This strategy of analyzing the notion of wanting was first suggested in an article by Richard Brandt and Jaegwon Kim [38], but it has since been taken up by Alston [2] and [3], Audi [8] and Tuomela ([205], pp. 77-80). I shall here concentrate on Audi, whose exposition of the analysis seems the most thoroughgoing. Audi proposes that the meaning of "xWp" ("x wants p") is fixed by the following set of laws:

- "W₁ If x does or would find daydreaming about p pleasant, and x believes that there is at least some considerable probability of p's occurring, then, under favorable conditions, xWp.
- W₂ If (a) it would be (or is) unpleasant to x to entertain the thought that p is not (or probably is not) the case, or the thought that p will not be (or probably will not be) the case, and (b) x believes that there is at least some considerable probability of p's occurring, then, under favorable conditions, xWp.
- W₃ If xWp, then (a) x tends to think (reflect, muse, or the like) or daydream about p at least occasionally, and especially in idle moments; and (b) in free conversation x tends to talk at least occasionally about p and subjects he believes to be connected with p.
- W₄ xWp if and only if: for any action or activity A which x has the ability and the opportunity to perform, if x believes either (a) that his doing A is necessary to p, or (b) that his doing A would have at least some considerable probability of leading to p, or (c) that his doing A would have at least some considerable probability of constituting attainment of p, then x has a tendency to do A.
- W₅ xWp if and only if: given that x has not been expecting p, if x now suddenly judged (or suddenly discovered, realized, or the like) that p is (is becoming, will become) the case, then x would immediately tend to be pleased, joyful, or relieved about this.
- W₆ xWp if and only if: given that x has been expecting p, if x now suddenly judged (or suddenly discovered, realized, or the like) that p would not be the case, then x would immediately tend to be disappointed about this.

W_7 xWp if and only if: under favorable conditions, x has a tendency to avow that he wants p ." ([8], pp. 4-5)

Though Audi maintains that each of these propositions expresses "part of what we mean by" (p. 10) " xWp ", he insists that they are, individually, empirically testable and that they together add up to an empirical theory (pp. 11-3, 15-6). I shall not quarrel with this, since it might well, I think, be true at least in the sense that it is conceivable that constructions of the form " xWp ", on different occasions of use, slide from one meaning to another, thus sometimes making one "law" analytic and others synthetic, but at other times making the former synthetic and one of the latter analytic. If there are multiple "criteria" for the application of a term, each one of the meaning-fixing rules may be empirically testable on some occasion, although there can be no single occasion on which all of them are simultaneously testable.

W_4 can be regarded as a rough counterpart to my definition of wanting - especially if conjugated with some material resembling the (a)-part of W_3 . So, when Audi remarks that W_4 is "probably the most important in the set" (p. 6), he shows that his position is not so far from mine. But still there is, of course, the difference that he also views W_1 - W_2 and W_5 - W_7 as partly explicative of the notion of wanting. My reason for not including anything like W_1 , W_2 , W_5 and W_6 in the analysis of wanting is that they refer to emotions or feelings - of pleasure, displeasure, disappointment etc. -, and the reasons for making the meaning of "wanting" dependent on that of feeling-terms are not stronger than the reasons for reversing the dependency. Audi is, however, willing to grant that there is a reciprocal dependency between the notions of having an emotion or feeling and having a want (pp. 14-6), arguing that this circularity does not imply uninformativeness, because there are also other ways of fixing the meaning of "having a want" and "having an emotion", other explicative laws - like perhaps W_3 and W_4 in the case of wanting. But this just reinforces my point that it is only through W_3 and W_4 that "wanting" gets its distinctive meaning; it is hard to see what contribution W_1 , W_2 , W_5 and W_6 can make to its meaning, given that the mentioned interdependency holds.

Nor does the inclusion of W_7 seem to enrich the content of "wanting", for it presupposes that "wanting" has been assigned a meaning beforehand, otherwise the awoval would be meaningless (cf. [3], p. 331); besides, as Audi recognizes ([8], note 5), non-language users can perfectly well have desires, and it is gratuitous to suppose that "desire" here means something else. Thus, I would conclude that the full meaning of "wanting" is set out by conditions

resembling W_4 and the first part of W_3 .

Let me argue in another way against Audi's thesis that a necessary condition for fully grasping the ordinary notion of wanting is to know W_1 , W_2 , W_5 and W_6 or, more generally, to possess the concept of emotion. First of all, W_5 and W_6 are not true unless a qualification is attached to them: xWp is sufficient under the stated conditions for x 's tending to be disappointed or pleased only if it is added to the conditions that x has the capacity to feel emotions at all. It seems clear that this capacity cannot be contained in the capacity to have wants, for the capacity to have emotions is a capacity to feel, and wanting does not necessarily include feeling anything (cf. e.g. [203], p. 53). What is felt in experiencing an emotion is, I suggest, those involuntary bodily disturbances - e.g. quickened heart-beat, pilo-erection, dryness of throat, visceral agitation - which characterize different emotions.

Now imagine that there is a species of creatures which is in every respect like human beings, save that the efferent pathways, which are necessary for the production of the involuntary bodily reactions, and/or the afferent pathways, essential for the feeling of these responses, are lacking. These organisms will as a consequence be incapable of experiencing emotions (except possibly in a very attenuated sense). Therefore, if they do not have any social intercourse with creatures capable of emotions, their language will be in want of emotion-words, which will make it impossible for them to frame propositions like W_1 , W_2 , W_5 and W_6 . But there seems to be no reason to deny that these organisms are able to act intentionally - for one thing, the efferent pathways between brain and the so-called voluntary muscles are, ex hypothesi, intact. Of course, this entails that they are capable of having wants as well. Furthermore, their language will in all likelihood contain turns of phrase which will express this aspect of their experience. But if so, it is possible for them to comprehend the meaning of "wanting" without having any knowledge of W_1 , W_2 , W_5 and W_6 , without having any understanding of the emotional vocabulary. This seems to show the independence of the concept of wanting vis-à-vis that of having an emotion.

It might be objected (cf. Audi, pp. 8-9, 11) against this that W_1 , W_2 , W_5 and W_6 are obviously not purely synthetical: surely, if we e.g. exchanged "pleased" etc. in W_5 and "disappointed" in W_6 for each other, we would get not merely a falsehood, but a conceptual monstrosity. Such an objection would, however, be based on the illicit inference that since having a want and having an emotion are not to be defined in terms of each other, they cannot be analytically related either. To see the invalidity of this inference, consider the following simple illustration: "brother" and "sister" need

not be defined in terms of each other, yet e.g. the conditional "If a is a brother of b, then, if b is female, b is a sister of a" expresses a logical truth. I will now try to explain - but for reasons of space only in a very sketchy fashion - how the analytical relationships between wanting and having an emotion are to be construed.

It will be convenient to conduct the discussion in terms of the negative attitudes of aversion for p (wanting not to bring about p or wanting to bring about -p) and negative emotions like fear, disappointment, sadness etc. that p. The two facts to be explained can then be formulated:

- (A) why it is necessary that if a is averse to p, then if he instead had had an emotion directed towards p, it would have been a negative one, and vice versa that if he has a negative emotion oriented towards p, then if he instead had had an attitude of wanting directed towards p, it would have been aversion;
- (B) what determines whether a's attitude towards p, granted that it is to be negative, will be one of aversion or one of negative emotion.

Consider a type of proposition that can clearly be the object of both aversion and negative emotion - the type which involves the idea of oneself experiencing a sensation of pain. In the present tense case, where I judge that I presently feel pain, one can divide the elicited reflex-reactions into three categories. (1) There is the reaction of withdrawing or shying away from the source of pain - a reaction whose obvious function is to terminate the sensation (and thus minimize the damage that is its cause). (2) There are reactions which, like those under (1), involve parts of the body that are voluntarily controlled, but which are not designed to terminate the pain-sensation, e.g. screaming, grimacing, tensing one's body, driving one's nails into the palm of one's hand. The function of these may also be to mitigate the experience of pain, but if so, not by eliminating the cause of the sensation, but by creating alternative sense-impressions, which may draw one's attention from the pain-sensations. (3) There are reactions which concern parts of the body that are outside the direct control of the will, e.g. sweating, trembling, quickened heart-beat.

A proposition involving the idea of one's feeling pain can be the object of aversion only to the extent that this proposition represents this idea as something that one might escape altogether or at least might get rid of. Therefore, one can view aversion as an extension of the form of reaction under (1), as the primitive drive to escape or eliminate pain being channelled from simple automatic behaviour patterns to behaviour consciously calculated to achieve this end. I now suggest that a proposition involving the idea of

one's feeling pain can be the object of negative emotion only in so far as this proposition represents this idea as something that one might not be able to avoid or even terminate. Which negative emotion one experiences is, in some measure, dependent on how certain one is of one's own powerlessness: if one is absolutely sure, one will feel (e.g.) despondency or despair; if one is less than quite sure, one will feel (e.g.) fear. (Thus, it is conceivable that the same content is the object of both aversion and fear, since that one might be able to escape a source of pain has a complement in that one might not.) Because emotions are evoked by ideas which one sees oneself as incapable of turning into reality, the emotional reaction can include items under (2) and (3) - and sensations thereof. These are reactions which are not in any way designed to escape or eliminate the sensation. But the reactions under (1) are not excluded either, since, as we saw in the torture-example on p. 106, they might survive even where one is convinced of their futility, though they cannot here be elevated to the rank of manifestations of intensional wanting, i.e. thought-guided behaviour (since thinking that what one is causing is p requires thinking that one can cause p).

The outlines of my explanation of (A) and (B) can now be discerned. Ad (A): it is necessary that the objects of aversion and negative emotions contain a common conceptual component in their objects, since there is an overlapping in the reactions (caused by the object) they involve, namely in the reaction of shying away, which in the case of aversion is transformed into thought-guided behaviour to achieve this end. Ad (B): what determines whether this common conceptual core gives rise to aversion or negative emotion is whether it is represented by the agent as something attainable or unattainable by his efforts. Since the object of wanting is something regarded as within one's power, the reactions under (2) and (3) - which constitute the major part of what is felt in emotion - do not develop.

Audi would not be pleased with the conclusion that my argument has led up to - that is, that W_1 , W_2 , W_5 and W_6 , though not explicative of the meaning of "wanting" are clear-cut analytic truths. He argues against the thesis that "the explicative set for wanting contains only or even primarily analytic propositions" (p. 13) by drawing some logical consequences from the set, consequences which he deems to be obviously non-analytic. Now I am of course not compelled to embrace the analyticity of all these consequences, since I reject the analyticity of some propositions in the set - e.g. W_7 - and regard others - especially W_3 - as serious mis-statements that contain only a grain of logical truth. But if my argument has been correct, it shows that the following is a (somewhat inaccurately formulated) logical truth:

"(B) If x would immediately tend to be disappointed to discover that $\bar{p}$ would not be the case, then, if x has the ability and opportunity to do A , and believes that his doing A has at least some considerable probability of leading to $\bar{p}$, x has a tendency to do A . (From W_6 and W_4)" ([8], p. 14).

My overall contention is that nothing has been unearthed that indicates the untenability of the assumption that the concept of wanting can be defined by means of a set of logically necessary and sufficient conditions.

6. THE CONCEPT OF A REASON-GOVERNED ACTION

6.1 The Problem of How Reasons Govern Actions.

1. The point of insisting that the characterization of a proposition as a reason for action involves a reference to attitudes of wanting, is that this should make it intelligible how consideration of reasons can animate the reasoner, can move him to act at will. After having, in the last chapter, defended the existence of this conceptual connection between the notion of a reason for action and the notion of a want and after having clarified the latter notion, it is now time to see whether these accounts allow a plausible construal of how reasons govern actions. By a "reason-governed action" I mean an action that is performed because of and in accordance with a reason, in such a sense that both the antecedent and, by it, the consequent of the reason-constituting conditional are intentionally made true, i.e. p and by it q are both intentionally made true if the reason is that if one brings about p , q .

The task of elucidating the concept of a reason-governed or reason-guided action falls into two parts. First, there is the problem of explaining how wanting q and thinking that (only) if one brings about p is q brought about, on the one hand, connect with wanting p , on the other, i.e. the problem of the transferral, in practical reasoning, of attitudes of wanting towards states of affairs basically causable. Secondly, an account must also be given of the transition from a want to an action executed at will. The latter problem could in its turn be subdivided into two: (i) the problem of the voluntariness of basic actions, and (ii) the problem of the voluntariness of non-basic actions. A necessary condition of the intentionality of a non-basic action is presumably the intentionality of the basic action that generates it: if a acts for the (phenomenal) reason that, as he thinks, if he causes p then q will be caused, he can only intentionally cause q if he intentionally causes p . Hence, the problem of the intentionality of basic actions must be solved before the corresponding problem about non-basic actions stands any chance of being solved.

I think the first problem - about the transferral of wants in practical reasoning - can be dealt with comparatively quickly, against the background of my analysis of attitudes of wanting. If a , at t , thinks that if he causes p then q will be caused as well, and he wants _{d} q , then it is by no means necessary that he, at t , wants _{d} p : a person is clearly not compelled to want everything he regards as a sufficient means for what he wants. The set of sufficient means may not be consistent, and even if it is, the agent may have

strong aversions against some sufficient means, either for their own sakes or because of other consequences of them than q . However, he must want each and every sufficient means in so far as they lead to q , i.e. if he were to base his attitude towards them solely on the fact that they are sufficient measures for q .

What is necessary, further, is that a wants_d to apply some means or other he considers sufficient for his bringing about of q . Thus, if he, at t , holds that it is true both that if he brings about p then q and that if he brings about r then q , but that there is no other sufficient means for q , then he, at t , must want_d either p or r . (Which of these a wants_d depends on the strength of his attitudes of wanting towards p and r for their own sakes or for their other consequences.) Another way of putting the same point is: an agent must want_d whatever he considers a necessary measure for something he wants_d; for here causing either p or r is thought to be necessary for accomplishing q . The "must" here is a logical must: it is logically impossible that a , at t , wants_d q and thinks that only if he brings about p then q , but does not, at t , wants_d p ,¹ for, as we saw in 5.4.2, it is part of the concept of wanting something, held to be realizable only indirectly, that one tends to calculate the means of achieving it and put the necessary means into practice when found. Thus, if a , upon discovering that p on the occasion at hand is a necessary (and sufficient) means for q , finds himself not prepared to want_d p at the time, he can no longer be said to want_d q at the time (though he might still want_r q , if p is not in the normal course of affairs necessary for q , or if p in this particular situation has atypical undesirable consequences).

This account of practical reasoning parallels that given of theoretical reasoning in 3.2.2, where it was said that it is logically necessary that if a , at t , thinks that if p then q and that p , he must, at t , tend to think that q . In both cases, readiness to draw the conclusion, to want p and to think q , respectively, is not something that is required of one given that one is to be rational (conceived as something that is a matter of choice), but something that is logically necessitated by the attribution of the former thought-complex or thought-want-complex, something whose falsity precludes the attribution of these complexes.

In order to strengthen the case for this interpretation of practical reasoning, let me critically review some alternatives to it.

Roy Edgley agrees with me that "it is logically possible to believe what is inconsistent only if it is not known or believed to be inconsistent".² But according to him, the situation is quite different in the practical sphere:

"reason may fail to decide his conduct in this way too, that he may not do y even though he intends to do x and thinks or knows that to do x he must do y" ([75], p. 144).³

In other words, it is said to be possible that a person wants p, realizes that only if he brings about q then p, but still does not want q. The realization that something is necessary for something wanted only creates a normative "pressure" ([75], p. 134), a pressure which it is possible consciously to resist (whereas in the theoretical field only non-conscious "resistance" is feasible). I take Edgley to be saying that this resistance to reasons is what we find under the rubric "weakness of will" (pp. 122, 144) and that if this phenomenon were not possible, there would be no "concept of the will" (p. 122).

This view of reasons in the practical dimension gives rise to certain troubles for one of the central tenets of Edgley's book, to wit that "a reason is a fact that bears normatively on what it is a reason for and can explain someone's 'doing' what it is a reason for 'doing'" (p. 154). Otherwise expressed, in the notion of a reason - be it theoretical or practical - a normative and an explanatory function are welded. In the theoretical sphere, the interplay between these two functions is not difficult to comprehend: since it is logically impossible to think both that p and that (only) if p then q, but not to tend to think that q, it is easy to understand how reference to the occurrence of the former two thoughts could explain the occurrence of the third. In short, it is understandable how the explanatory and normative functions can be combined here, for normativity must here be taken in Edgley's weak sense, i.e. as not requiring conscious defiance (see note 13 to chap. 3). But in the practical sphere where, according to Edgley, the connection between wanting p and thinking that p only if q, on the one hand, and wanting q, on the other, is normative in the stronger sense of conscious violation of the prescriptions of reason being ontically possible - how could reference to the former two attitudes be enough to explain the presence of the want that q? In order for the latter to be fully explained, it appears necessary to add a clause which excludes that the subject falls victim to akrasia, weakness of will (as conceived by Edgley). It seems a complete mystery how the same conditions can function both explanatorily and normatively (in my strong sense), for the former role implies, in the terms of my example, that not wanting q is ontically impossible (given wanting p and thinking that only if q then p), while the latter function implies that not wanting q is ontically possible (given the same). If thinking that p only if q and wanting p bear normatively (in the strong sense so that weakness is possible) on wanting q,

then surely the former cannot by itself explain why the want that q occurred. So far as I can see, Edgley does nothing to dispell this mystery.

David Richards [170] proposes a theory of practical rationality that is normative (in my strong sense) just like Edgley's theory of practical reason. Richards distinguishes between "principles of rationality", the existence and applicability of which constitute rational reasons, and "principles of morality", the existence and applicability of which constitute moral reasons (see [170], chaps. 2 & 4). Only the former principles are of interest here. It is clear from the way Richards formulates these principles that he takes them to be normative in the sense that it is possible consciously to defy them. For instance, one principle is formulated thus:

"Effective means: given a desired end, one is to choose that action which most effectively, and, at least cost, attains that end, ceteris paribus" ([170], p. 28).

The "is to" idiom is obviously as a rule normative. To remove any doubt about Richards' use of it: "the agent may choose to act or not to act on rationality principles, as he wishes" (p. 51). Richards adds his opinion on the way this choice commonly goes: "as a brute fact of human psychology, there is a widespread desire to be rational" (p. 63).

Could this desire to be rational, which Richards assumes, help us with the puzzle Edgley left us with? More concretely, is this desire to be rational what has to be added to the want that p and the thought that p only if q in order for a sufficient condition for the want that q to arise? There are at least two objections to such a view.

First, this desire to be rational that Richards generously ascribes to human beings cannot be founded on phenomenal reasons. Nor can it be justified in terms of real reasons, that is, the question "Why should one desire to be rational?", interpreted as "What real reasons are there for one to ...", must be ruled out as illegitimate, although desiring not to be rational is an option. For otherwise one would be forced to postulate a set of higher order principles, the problems of which would duplicate those of the lower order principles, and so on ad infinitum (cf. 5.1). So, the desire to be rational must itself be non-rational, i.e. logically beyond justification in terms of reasons. This seems to be a disquieting conclusion.

Secondly, appeal to a desire to be rational can not really complete the above explanatory sketch. For if "weakness" can afflict the want that p, why could it not afflict the want to be rational as well? That one desires to be rational clearly does not, on this account, entail that one succeeds in being rational (because of the possibility of "weakness"), and apparently only the addition of the latter will make the explanatory condition sufficient for

wanting that q . Hence, it must be added that the desire to be rational itself succeeds in being rational. But if this is conceded, why not instead append a clause demanding that the want that p is rational? That would have the same effect of making the explanatory condition sufficient; so the introduction of a desire to be rational is quite superfluous.⁴ A deeper trouble, however, is that this appendage consisting of a clause demanding that a want (be it a want that p or to be rational) is rational on a particular occasion does not appear to be genuinely explanatory. For what does it mean that someone's want that p behaves rationally, on an occasion where he thinks that p only if q , if not merely that he also wanted q on that occasion? But what is to be explained is precisely why this is so, why a want that q resulted in these circumstances, though, ex hypothesi, it does not result in all circumstances which include a want that p and a thought that p only if q .

So, the problem of how reference to a want and a thought of these last mentioned types can explain a want that q on a normative account of reasons still remains unsolved. A possible suggestion is this: a special neurological factor, which has no psychological equivalent, is present when and only when the want that q results. But, so far as I know, current neuro-physiology provides no evidence of such a factor. Further, this idea seriously disturbs the intuition that explanations in terms of wants and thoughts possess self-sufficiency.

Another stand on this issue is taken by Thomas Nagel [144]. Nagel claims (a) that the belief that one has a sufficient reason for acting "can sometimes fail to prompt action or desire without any explanation" ([144], p. 65, my emphasis), and that one cannot go further than asserting that, when no "contrary influences" - such as weakness of will, laziness, panic etc. (p. 66) - are operative, it "can by itself motivate ... to the appropriate act or desire, and usually does" (*ibid.*). But he still holds (b) that "the acceptance of such a judgement is by itself sufficient to explain action or desire in accordance with it, although it is also compatible with the non-occurrence of such an action or desire" (p. 109). In short, Nagel would say about the explanatory conditions cited above that they neither constitute a sufficient condition for the want that q , nor can be augmented into one, but that they still make up an adequate explanation of sorts, presumably a probabilistic one.

Nagel's view is particularly exposed to an objection which, however, can be rephrased to tell against any view which assumes some contingent relationship between wanting_d p and thinking that p only if q , on the one hand, and wanting_d q , on the other (even the view that the former is causally sufficient

for the latter). If sometimes, inexplicably, the former complex of attitudes is not accompanied by wanting_d q but with not wanting_d q, this could at any moment become the rule; perhaps from now on people will regularly be averse_d to q when they want_d p and think that p only if q and claim that this constitutes rationality. But this is strongly counter-intuitive: surely it is not just a lucky coincidence that the want_d to cause q rather than the want_d not to cause q shows up in this context.

On my view this is no mere lucky coincidence, nor is it due to people being wired up in a certain way, to their possessing a particular neurological make-up. If what, to all appearances, was a want_d not to cause q, regularly co-existed with what, to all appearances, was a want_d to cause p and a thought that p only if q, that would merely show that appearances were deceptive. A re-interpretation of some attitude would be necessary: perhaps we are not dealing with wants_d (though the subjects themselves may be using the term "dominating want"), or perhaps the content of their thought cannot properly be rendered with "only if", as we use it (though again the subjects' reports may be couched in these terms). It is logically impossible simultaneously to want_d to cause p, to want_d not to cause q and to think only if q, p; wanting_d not to cause q in the light of the thought p only if q is tantamount to having given up the want_d that p (if one ever had it).

2. Let us now turn to the other problems, those of specifying what relation must obtain between a want and an action for the latter to be rendered intentional. Beardsley [27] has returned a very simple answer to this question, namely one roughly consisting in that the want and the action be simultaneous and that the latter be in accordance with the former, i.e. that it realize the desired object at the desired time. More precisely:

"Z did D intentionally at t iff

- (a) Z did D at t,
- (b) Z knew-at-t that he was doing-at-t D, and
- (c) Z wanted-at-t to do-at-t D" ([27], p. 166).⁵

So, Beardsley is not demanding that the want (or the knowledge) plays a role in the production or explanation of the doing.

A somewhat fanciful example will show that Beardsley has failed to state a sufficient condition. Suppose that a surgeon has, unbeknownst to a, cut off the efferent pathway leading from a's brain to his right arm. a is still at t unaware that he has been incapacitated thus. At this time, he wants to raise his right arm instantly. Simultaneously, the surgeon electrically stimulates the end of the pathway he has severed from its connection to a's brain, thereby causing a series of muscular changes whose upshot is that a's

right arm goes up. Since a will in all probability be aware of his arm going up (due to muscular contractions in the arm), all conditions of Beardsley's analysis will be satisfied. Still, there can be no doubt that a has not intentionally raised his arm at t; indeed, it is not even true that a has performed the action of raising his arm - it seems more apt to say that the surgeon has raised it, non-basically, to be sure, by stimulating the nerve-endings.⁶

This example has been directed specifically against Beardsley: it hinges on (in this context) inessential features of his doctrine (namely, his concept of a deed or doing). But I trust that with the help of it, and the further illustrations to follow, the reader will be able to construct for himself counter-instances to every other variant of the type of theory that denies the necessity of demanding for the appropriate want a role in the explanans of the action to be made intentional. Hence, I take it as established that wants, perhaps together with thoughts constituting a phenomenal reason, must figure in the explanation of an action done at will.

However, it will not do simply to say, that the action must occur because of the appropriate wants and beliefs. In order to make this apparent and to sharpen our awareness of the intricacy of the issue, let me describe a difficulty posed by Davidson⁷ for the view that an action is intentional if it is caused by reasons that (as he expresses it) rationalize the action, i.e. that the action of bringing it about that p is intentional if it is caused by a thought or a belief that if p is brought about then q is brought about as well and the want that q. The hitch is that the reason may be linked to the action via "non-standard or lunatic internal causal chains" ([68], p. 153). Davidson provides the following illustration:

"A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally" ([68], pp. 153-4).

Davidson anticipates the temptation to get round the difficulty by taking a certain route and expresses his suspicion that the route is blocked:

"It will not help, I think, to add that the belief and the want /described in the example/ must combine to cause him to want to loosen his hold, for there will remain the two questions how the belief and the want caused the second want, and how wanting to loosen his hold caused him to loosen his hold" ([68], p. 154).

I claim to have sketched the solution to the first of the problem that here agitate Davidson. Davidson goes wrong in thinking that the relation between wanting to get rid of the weight and danger of the companion and believing that if one loosens one's hold on the rope then one will achieve

this, on the one hand, and wanting to loosen one's hold on the rope, on the other, is causal. In fact, the relation is logical: a cannot correctly be described as wanting_d q in a context where he believes that p is the only sufficient means for q available to him, but still does not want_d p, because wanting q includes a tendency to cause what one thinks is q, and in the context of the described belief, this will be a tendency to cause what one thinks is p and by this to achieve q. So, on the assumption that the mountaineer's beliefs concerning means have been exhaustively specified, his wanting to loosen his hold follows logically. Davidson's (p. 151) fear that the conclusion-want might not obtain because the reasoner "fails to put" the premise want and belief "together" is unfounded. To a person who fails to put these things together it is incorrect to attribute either thoughts of a hypothetical content or wants having contents to be realized indirectly.

Thus, I do not think that adopting the strategy Davidson tentatively rejects doubles the number of problems to be solved; but we still face the problem of spelling out the relation that must obtain between wanting to loosen one's hold and the (we assume) basic action of loosening it in order for the latter to be intentional. It is not enough to say that the action must occur because of the want, for this might still be true without the action being done at will, e.g. if the want causes a fit of nervousness which in its turn causes the loosening of the grip, the latter occurs non-intentionally. We are still confronted by the problem of lunatic internal causal chains.

A seemingly easy and natural way out of this difficulty - suggested e.g. by Thalberg ([201], pp. 57-62) - is to require that the loosening of one's grip must result directly from the want to loosen it, if the former is to occur intentionally. But this simple proposal is not satisfactory, for the result of a non-basic action - e.g. that the climber gets rid of his companion - is produced indirectly by the want to get rid of him, yet a non-basic action can be intentional. So, not every type of indirect determination of a result by a corresponding want excludes intentionality; only the interposition of some peculiar factors does. Which? Simply enumerating them - emotional disturbances, brain-haemorrhages ... - will not do, for this answer is an ad hoc hypothesis whose explanatory value is nil. A general characterization of what distinguishes phenomena that annihilate intentionality must be provided, if any insight into the nature of intentional action is to be gained.

Perhaps Thalberg hints at this distinguishing feature when he points to ([201], p. 59) the unexpectedness of the affective disturbance from the point of view of the agent. In any event, this is in essence the thesis that Richard

Foley is advocating. Foley asserts that if we require of an agent, in order for him

"to bring about his hand rising deliberately and basically, not only to will his hand rising and to have his willing cause the hand rising, but also to believe his willing will cause the hand rising in the manner it does, we will have a device to make the account sufficiently restrictive" ([81], pp. 64-5).

This would explain how there could be intentional non-basic actions: this happens when the results of an agent's basic actions have the consequences he expects. According to this definition, Foley would maintain - and actually does concerning a similar example ([81], pp. 65-6) - that if a climber (a) foresees that if he wants to loosen his hold, this will cause a nervous excitement, which in its turn will lead to the loosening of his hold, but (b), in the face of this belief, still begins to want to loosen his hold, and as a consequence initiates the anticipated causal sequence ending in his relaxing his grip, then he will deliberately have relaxed his grip on the rope.

However, this seems to be an analysis of what it is to knowingly or wittingly (by willing) cause oneself to perform the action of loosening one's hold, rather than of what it is to perform the latter action deliberately. If I stick a finger into a flame, well knowing that this will cause me to pull it out on reflex, then I have knowingly and perhaps intentionally caused myself to perform the action of pulling my finger out of the flame; but it does not follow that I have knowingly and perhaps intentionally pulled the finger out of the flame: this remains a reflex, despite the fact that it has been predicted. We must not confuse the causing of a performance of an action with the performance of the action itself. My suggestion is that Foley focuses on the former notion instead of the latter, that he studies the causing of an action, i.e. the causing of a certain causal sequence (here consisting in the excitement and the loosening of the grip), and that he demands of this causing that it must be planned beforehand. But this has no bearing on the status of the caused action. If excitement is involved, the action appears to be just as involuntary as the action of pulling the finger out of the flame. As Foley himself notes (p. 66), once the agent has caused the excitement, he is no longer in control of what results - just as I am no longer in control of my movements once I have put my finger in the flame.⁸

At this juncture, defeatism can easily seize one: the efforts to specify the causal nature seem futile. I think that defeatism is the proper diagnosis of Goldman's attitude when he asserts that all a philosopher, qua philosopher, is capable of saying of the relation between the appropriate propositional attitudes and the basic action which is to be made intentional by their

presence, is that the former must cause the latter "in a certain characteristic way" ([88], pp. 57, 61-3, 166-9). Goldman delegates the task of spelling out in detail what this characteristic way amounts to to neurophysiologists.

However, there must be common-sense criteria for what counts as this characteristic way: this is evidenced by the fact that we are able to give verdicts in individual cases, e.g. the verdict that no intentional action was executed by Davidson's mountaineer. Furthermore, the neurophysiologist himself must presumably apply these criteria when he singles out what physiological phenomena to study: he must select those phenomena which occur when and only when an intentional action has been performed, i.e. just when the criteria for an intentional action are satisfied (cf. [215], p. 174). Perhaps Goldman does not disagree with this; at least he writes that "we can 'feel' a difference between a voluntary movement and an involuntary one" ([88], p. 62). This intuitive feeling is perhaps our criterion, in his view. But for this to be consistent with what was said about the impotency of philosophical analysis, this feeling must be held to be philosophically unanalyzable or indefinable.⁹ My complaint would then take the form of a complaint against an appeal to indefinables.

This survey of unsuccessful attempts to elucidate the causal tie alleged to hold between propositional attitudes and an action, if the latter is to be intentional, might encourage the view that these attempts are based on a radical misconception, namely that the tie is causal. Perhaps what counterexamples like Davidson's really demonstrate is the breakdown of the causal type of approach to intentional action, of the view that the intentionality of actions stems from the causal efficacy of the appropriate wants and beliefs. Frankfurt is a champion of this standpoint:

"causal theories are unavoidably vulnerable to such counterexamples /as Davidson's/, because they locate the distinctively essential features of action exclusively in states of affairs which may be past by the time the action is supposed to occur. This makes it impossible for them to give any account whatever of the most salient differentiating characteristic of action: during the time a person is performing an action he is necessarily in touch with the movements of his body in a certain way ...".¹⁰

Stoutland has likewise argued ([196], pp. 288-94) that Davidson's mountaineering-anecdote presents a decisive counter-example to a causal theory, that e.g. the directness required of the relation between the propositional attitudes and the action is not causal at all. Positively, Stoutland asserts ([196], p. 291) that a non-causal theory like von Wright's (see e.g. [217]) has no difficulty with a case like Davidson's. One crucial passage of Stoutland's reads:

"If it is true that an agent intends to get into the house, and believes that cutting a hole in the window now is necessary for that, then whatever his (mere) behavior may be, his intentional act is cutting a hole in the window (or trying to). That is what he intends (means) by his behavior, whatever it is, and that is how we must understand it. Whatever behavior is occurring, he is doing something intentionally." ([196], p. 280)

From this passage it may seem that Stoutland would just rule that the climber intentionally loosens his grip, because the loosening of the grip occurs in the context of a desire for whose satisfaction the climber believes that his loosening his grip is (to modify the example a little) necessary. But this is not so, for Stoutland demands that in order for the ensuing behaviour to be intentional, the desire must "become" an intention or the climber must acquire an intention to satisfy the desire ([196], pp. 292-3). It is, however, never made clear what this "transformation" of a desire into an intention amounts to. Surely, we cannot require that the desire is an intention only if it results in an intentional action - for that would be a circular condition in this connection. But if a desire can be characterized as an intention independently of the intentionality of accompanying behaviour - how can it then be true that this behaviour, whatever it is, must logically be a certain intentional act?¹¹ Even von Wright himself concedes ([219], p. 59) that it is "just possible" that the concurrent behaviour is not intentional, although he makes no attempt at an emendation that would exclude this possibility. So, it must be concluded that this non-causal approach has failed to solve our problem.

This could be taken as an indication that, in spite of our disheartening survey of causal approaches, we should be loath to give up this kind of approach. (A criticism of Frankfurt's non-causal theory in the next section will further buttress this standpoint.) For my own part, I cannot but believe that the causal approach represents the only hope of understanding the enigma of intentional action. But before embarking upon the project of re-instating a causal analysis, let me note two features that are conspicuously absent in Davidson's mountaineering-anecdote, but present in cases of intentional action. These should serve as tests of the adequacy of a causal theory.

One thing that is amiss is that the climber never tries to loosen his hold. As many writers have convincingly argued, acting intentionally entails trying.¹² In itself this observation does not take us very far, for trying is a notoriously obscure phenomenon, and so is its role in intentional action. Some have urged (e.g. [205], pp. 257-8) that tryings have to cause behaviour if an intentional action is to have taken place and that, more exactly, this is what is missing in the Davidsonian example. As we shall see, this, however,

is to misconstrue the concept of trying.

A second thing that obviously goes awry in the mountaineering-incident is that the bodily movements "go out of control", i.e. the bodily movements are not throughout their course dependent on the continuance of the climber's want to relax his grip: it is not the case that if he had stopped wanting this, the movements of his hands would immediately be accordingly modified.¹³ His weakness probably makes it impossible for him to tighten his hold.

If a causal analysis is to be vindicated, it must unite these two points about the initiation and maintenance of behaviour in a single account. Such a theory could well represent the literal truth behind Frankfurt's metaphor of the agent's being "in touch" with the movements of his body throughout the time of his acting. I will attempt to outline such an analysis in the next section.

6.2 Reason-governed Action: Basic Actions.

1. I shall start by giving a definition of an intentional basic action, for, as I remarked in 6.1.1, the intentionality of a non-basic action presupposes that of the basic action which generates it. Later on this definition will be extended to one of an intentional non-basic action. This will be seen to be precisely what defining the concept of a reason-governed action, as conceived in 6.1.1, amounts to. The notion of (non-)basicity involved here is not that of ontological basicity, but that of teleological basicity.¹⁴ That a brings about p in a teleologically basic action entails that there is no proposition q such that a thinks that he will make it true that p by making it true that q.

Here then is the promised definition:

(IBA) a brings about p during t- in an intentional (teleologically) basic action iff:

- (1) There is no q such that a, during t-, thinks that if he causes q during t-, p will also be caused by him during t-, and such that (2)-(5) remain true if "q" is substituted for "p";
- (2) Throughout t-, a wants_d to cause p during t-;
- (3) a's thinking, at t-, that he might be able to cause p during t- causally initiates what he, at t-, non-coincidentally thinks will (possibly) become p during t-;
- (4) a's thinking, throughout t-, that what he is (in the process of) causing during t- is (possibly) p, causally sustains, throughout t-, the perpetuation of that which a thinks is p;
- (5) The fact which a during t- thinks is p is really p.

Condition (2) might seem superfluous since what is of pertinence in it here is contained in (3) and (4). But I let it remain as a reminder of some of the findings in 5.4 that require no special stressing here, such as there being an intrinsic property of a's which is a precondition of the causal efficacy of the thoughts, and there possibly being reasons for a's wanting p.

Let me give a concrete illustration of (IBA): suppose p is those bodily movements that occur when one sits down, i.e. that one's knees bend in a particular fashion. Normally, there is no proposition such that an agent thinks that if he makes it true then he will have effected that his knees bend. For this set of bodily movements to be brought about at will, it must be at least an epistemic possibility for a, at the time just before the action commences, that he be able to bring about these movements. Furthermore, when this thought of ability begins to cause something to become a fact (perhaps because of the intrinsic attractiveness of bending the knees, but more likely because the agent associates further states of affairs with it), a must think that this will (possibly) be the bending of his knees. More than that, it must be the case that if a had not seen what is initiated as an instance of his knees bending, he (i.e. his ability-thought) would not have begun to cause it. It must not be a mere coincidence that a thinks of that which he begins to cause as bending his knees. Such is the purport of (3).

Now, as long as this causing goes on, a must keep to the thought that what he is in the process of causing is his knees bending. As a starts to act, to cause bodily movement, there will be a sensory feed-back, he will get sense-impressions - in this case proprioceptive and visual ones are of particular importance - of what bodily movements are in fact taking place. The requirement that a must keep to the original thought about what he is causing means that the judgements these sense-impressions license must bear out the thought formed beforehand. What happens when one discovers that an attempt to perform a basic action one is engaged in is unsuccessful - when one, for example, by looking into a mirror discovers that one's attempt to wiggle one's ears is unsuccessful - is precisely that subsequent sense-impressions falsify one's beforehand-thought about what will be caused or, in other words, that these sense-impressions justify judgements that are incompatible with the original thought. A reasonable conclusion might be that one, after all, lacks the ability in question. But the feed-back evidence, though disconfirmative, need not be as destructive as this: it may suggest how to adjust one's movements in order for them to conform to what is desired.

However, not only must a's thought that what he is causing is the bending

of his knees be present for the whole duration of this causing, it must all the time sustain the on-going of this pattern of movement, that is, at every moment during the occurrence of these movements it must be true that if a had abandoned this thought - perhaps in the light of feed-back information - his causing of the movements he in fact causes would have ceased. If this condition is not fulfilled, the whole action is not intentional, though its initial phases may be. Actually, sitting down provides a good example of an action whose later phases may not occur at will, though the earlier ones do, for when one has bent one's knees to a certain degree, one often loses one's balance and falls into the chair. When the movements reach this later phase, one has no longer any direct control over them: one can prevent them from following their course only by resorting to indirect measures, e.g. grasping something with one's hands. So much for the import of (4).

The last condition to be satisfied for it to be true that a bends his knees at will is that the state of affairs, which he causes to obtain because he, beforehand, thinks it will be that his knees bend, really turns out to be this. If this condition is not fulfilled, we have only an unsuccessful attempt to bend the knees. At least in most cases, to try to bring about p in a basic action is to bring about something you non-coincidentally think will be p. It is in agreement with the view of writers like Chisholm ([52], p. 53), Danto ([58], pp. 56-60) and Richard Taylor ([199], pp. 80-5), that when you try, you have to do something.

But should we really exclude a priori the possibility of instances of trying in which nothing is brought about at all? Consider somebody who regains consciousness after a traffic-accident: since he is not aware that the accident has made him totally paralyzed, it seems that he can try to move his body, although not the faintest muscular contraction occurs. Presumably, his eventual conviction that his body is paralyzed will be founded mainly on his having tried and failed. In order to make his theory of trying immune to this criticism, Taylor is willing to contend ([199], pp. 84-5) that one can have non-inferential knowledge of one's own capacities and incapacities, that e.g. the patient in our example is able to tell directly, without trying, that his body is paralyzed. So far as I can judge, this is a very desperate move, which can appear plausible only to someone who is determined to rescue a cherished theory. To my mind, it appears more reasonable to insist that there is after all something the patient (i.e. his false ability-thought) causes and to which his thought "A state of affairs which is now being caused to obtain is p" refers, e.g. something like electrical impulses leaving his brain and descending down his spinal column. Thus, I think it is possible to stick to

the doctrine that trying involves causing something.¹⁵

But my theory could also be stretched to cater for the case of the paralyzed patient in another way - a way which emphasizes, instead of effacing, the difference between this case and more normal instances of trying, where something is undoubtedly done. It might be said that so little is caused in the patient's case that the phrase "that which is now about to be caused" in the patient's thought cannot properly be held to refer to anything. It is not, as in more frequent cases of unsuccessful trying, where there is a referent, but where the predicative part is false of the referent. According to this suggestion, it would not in the former case be correct to employ the oratio obliqua construction "causes something which he thinks will be p". Having indicated this possibility of relaxing my account, I shall adhere to the first strategy, since it allows the conditions of (IBA), without modification, to do duty in an account of trying, as well. I take trying to bring about p in a basic action as being definable by (1)-(3) - plus (4) if one wants the trying to last through the doing.

From this it follows that intentionally causing p in a basic manner entails trying to cause p in a basic manner. In fact, the former is identical to successfully trying to cause p in a basic manner. A further corollary is that a trying is not a cause of that which is caused in the trying; it includes it as a part. For instance, a successful try to cause p is not a cause of p; p is incorporated as a component-event in the trying, since to try to cause p is to cause something, in the thought that it will be p, a something which in this case turns out to be p.

I hope nothing I have said implies that when a intentionally makes p true, this is the only proposition he makes true at the time (even in a basic manner). This would be obviously false: when a brings about p during t-, he is bound to make a host of other propositions true simultaneously, since what happens is a concrete event. When a bends his knees, he will inevitably bend them in a certain manner, at a certain speed, certain things will happen on the micro-level in his legs and so on. These are all things that he brings about non-intentionally - provided he thinks of what he is causing only as that his knees bend.¹⁶ (The cause of the manner of execution probably lies in certain behavioural habits of the agent.) Furthermore, we must remember that there are an indefinite number of non-synonymous phrases that can be used to refer to the individual pair of legs involved in a certain individual event: e.g. what a refers to as his legs might also be referred to as the stockiest legs in the neighbourhood or the only legs with such and such micro-structure in the vicinity. But this is not to say e.g. that a intentionally

bends the stockiest legs in the neighbourhood. For this to be true, "the stockiest legs in the neighbourhood" must be substitutable for "my legs" in the agent's thought, something which the norms of substitutivity laid down by (T6) in 3.2.1 do not permit. Considerations like these make it plain that the propositions an agent intentionally makes true at a certain time form only a tiny fraction of the total number he makes true at the time.

Can (IBA) cope with counter-instances of the Davidsonian kind? It should be able to do so, because it satisfies the two tests of adequacy mentioned in 6.1.2: it makes trying an essential factor in voluntary action, and it demands maintenance, by the agent's thoughts, of the bodily motion throughout its course or, in other words, it makes sustained trying an essential factor too. But let us look in more detail at the way (IBA) handles the Davidsonian counter-example. Suppose that the climber's thought that he is able to bring about that his hold on the rope loosens (and thus that he gets rid of the companion) causes a nervous weakening of his hands and consequently a loosening of their grip. This is not enough to satisfy (3), for it would be quite concordant with this situation to imagine that the mountaineer realizes that his hands are loosening their grip only after this has begun to happen, that he so to speak catches himself doing something he has already started doing. Thus, in a situation of the kind depicted, the agent may not concurrently view that which he is causing as his hands loosening their hold on the rope. But even if he does view it thus during the appropriate time-interval, this viewing is not normally an inus-condition of his grip's loosening, when this is caused by nervous excitement. So, there is more than one reason for denying that (3) has been satisfied, i.e. that the climber tried to loosen his hold in the act of loosening it.

This makes it clear that if we are to have a counter-instance against (IBA), the intervening factor must be inserted somewhere where it can intelligibly be assumed to be caused by the agent's "preview" of what he is bringing about - else the "preview" will be causally idle in the production of the result, and (3) will be violated. Suppose then that a's thinking that he is able to bring about p causes something which a non-coincidentally thinks of as p and that this something, due e.g. to excitement, causes p to materialize. Now, (3) no longer raises any difficulty, but instead (5) does, for if that which a thinks is p causes p, it cannot be p, and so (5) is not satisfied. It seems to me that wherever you insert an interfering causal factor, you will either violate the requirement that the agent's preview must be causally relevant - throughout the whole doing - for what is done, i.e. (3) or (4), or the requirement that this preview be true of what it refers to,

i.e. (5). Thus, I regard (IBA) as fool-proof against lunatic internal causal chains.

2. I would claim that (IBA) gives a satisfactory account of what Frankfurt calls "the most salient differentiating characteristic of action", to wit, the agent's being "in touch" with the movements of his body when acting. Frankfurt states that

"The performance of an action is ... a complex event, which is comprised by a bodily movement and by whatever state of affairs or activity constitutes the agent's guidance of it" ([83], p. 159).

I subscribe to a similar idea, viz. that an intentional basic action is the complex event described by (1)-(5). The difference between us is that I take the concept of guidance of or being in touch with bodily movements to be a straightforward causal notion, while Frankfurt resists this interpretation. His resistance seems to stem mainly from his belief that whereas causes precede their effects in time, "an event cannot be guided through the course of its occurrence at a temporal distance" ([83], p. 158; cf. p. 160 and quotation in 6.1.2).

There are two mistakes about the causal approach entangled in this remark. First, Frankfurt assumes (see esp. quotation in 6.1.2) that the causal theorist must locate the distinctive marks of an (intentional) action in the causes of the action itself, which implies in something that temporally precedes the action. But this is by no means necessary: to me, it is the causes of bodily movements - or other types of intrinsic results - that settle the question whether an (intentional) action has occurred. An (intentional) action is a complex event which includes these causing events (and their results). Hence, Frankfurt is mistaken when he supposes that the causal theorist must "locate the distinctively essential features of action exclusively in states of affairs which may be past by the time the action is supposed to occur" (p. 158). Secondly, the result of an action is a temporally extended process or sequence; and although this whole sequence is preceded by thoughts that initiate it in the first place, later phases of this sequence may be causally influenced by thoughts occurring simultaneously with - or even later than - the earlier phases of this process. This much should be apparent from my discussion of (4), of the importance of the agent's preview about what he is bringing about being borne out by sensory feed-back information. So, it is not even true without qualification that the result of an intentional action is preceded by those guiding thoughts the presence of which is logically required for an intentional action to have taken place. I conclude, therefore, that Frankfurt's argument against a causal analysis

of the concept of guidance breaks down here.

Frankfurt's own positive account of guidance (p. 160) is that bodily movements proceed under guidance if there are causal mechanisms, which stand ready to affect and adjust the course of the movements should the feed-back information indicate that the movements are on the wrong track. Thus, he does not demand actual causal influence on the movements, but only a potential influence that is triggered off when something is observed to go awry. The weakness of this is that it only explains why a sequence of movements, once initiated, is rectified (if it is); it does explain what set it off along its original path. Reference to desires or intentions or whatever as initiating causes could explain this only if it is taken to include or to be supplemented with a clause saying that the agent sees what he brings about as gratifying his desire. For a desire to cause p can explain why p is intentionally caused only if it is implicitly assumed that the agent correctly conceived what he caused as p. But if Frankfurt concedes this, he has surrendered to a causal approach after my model. Thus, I conclude that Frankfurt's positive account is not so much false as gappy and that its deficiencies can be filled only by a full-blown causal theory.

The core of my conception of an intentional basic action is this: when a brings about p at will, in a basic action, a must have a true thought about what he is about to bring about, i.e. he must think it will be p, and this thought must enter into the explanans of the materialization of p. If anybody feels tempted to dispute the presence of such a thought among the causal antecedents of the intrinsic result, I offer for his edification the following striking observation by William James. James reports ([107], p. 105; cf. p. 490) that when people, who suffer from anaesthesia in their arms, are asked to close their eyes and thereafter to raise their arms - which, unbeknownst to them, they are prevented from doing by someone's taking hold of their arms - they are surprised, upon opening their eyes, to find that their arms have not gone up. If they are astonished to find this, they must have thought the contrary, that their arms had gone up.

How is the presence of this thought to be explained? In the ordinary case of people intentionally moving their bodies, it is tempting to explain their thoughts about what they do as altogether the upshot of (outer or inner) perception of their movements, since the thoughts generally correspond to the movements caused. But in James' case this sort of explanation is impossible, because the subjects neither see nor feel. Now given the implausibility of the hypothesis that the presence of the thoughts is accidental, the order of dependence must rather be the reverse: what movements do occur is dependent

on what is thought to occur rather than that, as in the case of observation, what is thought to occur depends on what does occur. The subjects' false beliefs - that they can raise their arms - cause what they wrongly take to be their arms rising. Normally, people are correct in their beliefs about what they can do with their bodies, so what they think is their arms rising is indeed this; hence, their thoughts can be explained as arising from observation, sensory feed-back. Or, if they are wrong, they are quickly corrected by the sensory feed-back from the initiated events. The delusion lingers in James' subjects because the feed-back routes are blocked - hence, its usefulness for the purpose of demonstrating the existence of observation-independent thought about what is caused.

To some extent, this central idea of movements being caused because they are beforehand conceived to be of a certain kind seems foreshadowed by Anscombe's doctrine of "practical knowledge": she appears to claim that an agent has "non-observational" knowledge of his intentional actions and that this knowledge determines its object, i.e. what is done.¹⁷ But there are important differences between Anscombe's view and mine. First, she seems to classify knowledge based on proprioception, in contrast to knowledge based on outer perception, as "non-observational". Consequently, she regards one's knowledge that one's leg is moving in the patellar-reflex as "non-observational", when it is not based on outer perception, like e.g. seeing ([4], §§ 8, 28). I find this segregation groundless, for knowledge founded on proprioception is like knowledge grounded on outer perception, in that both are determined by movements occurring and so cannot be determinants of these very movements.

A second and more significant difference is that Anscombe sets up ([4], § 48) a dichotomy between "speculative" or "contemplative" knowledge, the content of which is determined by what is the case, and practical knowledge, the content of which determines what is to be the case. Now, there might well be an innocuous interpretation of this duality, according to which it is both genuine and illuminating. But it is untenable, if it is taken to imply the perplexing thesis that there is a propositional attitude that can correctly be labelled "knowledge", not because its content is formed to fit an independent reality, but because, regardless of any such prior fit (theoretical truth), it is sufficient to ensure correspondence between its content and the world by having an impact upon the latter.

Let me expand on this point. An agent's thought that he is able to bring about p during t- is a thought of the contemplative (theoretical) species - commonly, it is based on an inductive inference from the agent's experience

of having successfully brought about p on other occasions which appear similar in relevant respects to the present one. This thought is hypothesized (by me) to cause what the agent thinks is p. Now the agent's thinking of something that it is p does not, of course, turn it into p - witness e.g. James' patients. Whether or not what the agent thinks is p actually is p is wholly dependent on the truth - i.e. theoretical truth, correspondence with reality - of his thought that he can bring about p at the time. Thus the question of whether the agent's preview of what is done is true boils down to a theoretical issue: the truth-value of the preview is the same as the (theoretical) truth-value of the ability-thought; there is no question of the preview by itself guaranteeing that the world fits its content.

Kent Bach has evolved an analysis very similar to the one expressed in this section:

"I conceive of action as involving cycles of effective representations, bits of behavior, and receptive representations, in which there is a causal relation between each effective representation and bit of behavior, each bit of behavior and receptive representation, and each receptive representation and the next effective representation" ([19], p. 366).

Bach's effective representations are obviously parallel to my conception of the agent's preview of what he is about to cause in that both are assumed to be causal determinants of what is caused. His receptive representations share, with my construction of the agent's observation-based thought about what he has caused, the feature of being determined by what is caused. The effect of receptive representations on effective ones corresponds to what I have spoken of as the agent's preview being either corroborated or contradicted by perception-grounded thoughts. Therefore, I can confidently direct the reader to Bach's paper for further illustration of the theory propounded here - but not without indicating some significant points of difference between Bach's treatment and mine.

Representations, as conceived by Bach (e.g. p. 368), are sensuous, not linguistic; in the case of bodily movements they are generally of the kinaesthetic kind. Receptive representations are - in contrast to their equivalents in my theory, perception-based thoughts - simply identified with perceptions (*ibid.*). Effective representations are held to be closely similar to imaginings, in fact so similar that Bach eventually despairs of finding any informative way of characterizing the difference and tentatively concludes that they might after all be "imaginings of immediate changes" (p. 373).

Bach is absolutely right in claiming that the agent's thought beforehand about what is to be done (at least sometimes) cannot be linguistically coded,

that it must be couched in images. For even adult people are incapable of verbally describing in detail more than a tiny fraction of the repertoire of actions they can perform with their bodies: most people can e.g. swallow and walk, but who can describe in words what this is like?

But how is it they can conceive bodily actions in isomorphic symbols? They have perceived (especially felt) themselves do them; their symbols are derived from these perceptual experiences. This presupposes, however, that there is a class of actions the execution of which does not involve any pre-view on the part of the agent. I think there is in fact a class of such actions: the forms of reflex-behaviour that are innate in human beings and other animals. As I have repeatedly stressed, no thought about one's own ability to act - and hence no thought about what is being done - enters into the genesis of reflexes. Swallowing can serve as an illustration. The hypothesis is that an infant learns to swallow at will by deriving an image from its proprioceptive experience of swallowing on reflex, an image which it then applies in swallowing at will. On this hypothesis, acting at will is empirically posterior to acting on (innate) reflex.¹⁸

However, this picture of the acquisition of the capacity for thought-guided behaviour does not appear to square with everything Bach asserts. He puts forward his "representational theory", not (as I would do) as a theory about what is distinctive of acting at will, but as a theory about what it is to act in general, which includes acting "automatically, routinely, and/or unthinkingly" ([19], p. 363; cf. p. 367).¹⁹ (Since Bach regards even unthinking behaviour as occurring under the supervision of representations, he plainly cannot take representations to be thoughts. I find this odd: surely, to call up an image of something is to think of it.) But it seems that this leaves Bach with the problem of accounting for the genesis of the images figuring in effective representations: from the experience of what are they derived? Bach would be better off, I believe, if he narrowed the applicability of his "representational theory".

Another difference between Bach and myself is this: Bach denies (pp. 366, 368ff) that his representations are propositional in form. The proper linguistic rendering of his receptive representations is, he claims, by participial forms, "a F-ing" (e.g. "fingers moving"), whereas effective representations are properly expressed by infinitival constructions of the type "a to F" (e.g. "fingers to move"). It is the problem of the nature of effective representations that is especially acute for the action-theorist, since clarification of receptive representations is rather a task for the philosopher of perception. Bach concedes (pp. 371-3) that he is unable to pinpoint what it is

about the sensuous effective representations that makes the infinitival form appropriate and revealing. I understand his quandary. This linguistic form seems to suggest a peculiar practical or executive modality - it is reminiscent of the construction "a is to F".²⁰ But it is extremely difficult to envisage what possible feature of a sensuous representation could correspond to this linguistic modality.

I do not know to what extent Bach is tempted to postulate a peculiar practical, volitional or conative modality or setting in which the sensuous representation is embedded. His eventual concession that effective representations might just be "imaginings of immediate changes" and that the infinitival form might be a mistake, indicates that he does not find the temptation irresistible. The temptation to appeal to such a modality is probably at its strongest when one seeks the differentiae of intentional actions. For as we have seen, the image e.g. of oneself vomiting may cause one to vomit without the vomiting being intentional. So something is needed over and above an image of bodily change which is the cause of a suitable change for an intentional action to have occurred - perhaps that the image is embedded in a special practical setting.²¹

My theory of intentional action possesses the virtue of simplicity in that it eschews the idea of some causally potent, practical modality or copula being expressible in a natural, sensuous code. But - in opposition to Bach (and probably James as well) - the assumption that effective representation is propositional is made the heart of the theory: effective representations refer to what is in the process of being caused and attribute a property (isomorphically expressed) to it. That this assumption contains nothing superfluous or extravagant is indicated by this reflection: it seems that it provides the simplest imaginable explanation of something that is commonly held to be evident, namely that someone acting intentionally necessarily has a correct view of what he is doing.²² It is, of course, true that this theory is saddled with the problematic conception of a non-linguistic code, consisting at least partly of images, which is powerful enough to express propositions. But if my argument in 3.4 is correct, this is a conception which other considerations than action-theoretical ones force upon us. The postulation of a causally potent, practical modality would, however, be prompted exclusively by action-theoretical considerations.

There is a further respect in which my theory is superior to Bach's. Bach does not try to account for the motivational power his theory bestows upon e.g. kinaesthetic images of bodily change. Therefore, he comes to resemble James who relies on the principle "every representation of a movement awakens

in some degree the actual movement which is its object" ([107], p. 526). I find this principle dubitable; but even if it is true that images of bodily change have intrinsic motivational power, tend to cause the bodily changes which they depict, we still need to explain how motivational power can be channelled from one state of affairs to another. For most bodily changes are not just brought about for their own sakes, but for a reason, i.e. thoughts of states of affairs which are somehow linked to the bodily ones contribute to the motivation. By fitting my theory of thought-guided behaviour into a theory of reasons for action, I have filled this lacuna.

Annette Baier has stated that she does "not think it unreasonable to test a philosophy of action by applying its analysis to mental acts" ([22], p. 33). Now although (IBA) is designed primarily with a view to actions having physical results, I think it would stand up well to this test of adequacy. I am strongly inclined to believe that (1)-(3) and (5) state necessary and sufficient conditions for, say, a's intentionally visualizing a squirrel, i.e. his causing an image of a squirrel to occur in his mind in an intentional (teleologically) basic action. (4) is superfluous because the image is conjured up instantaneously. Generally, a condition like (4) is bound to cause trouble in cases of mental actions, since this monitoring or surveying of concurrently on-going mental processes - being itself a mental process - inevitably interrupts the course of the latter. For this reason - namely, that concurrent surveying equals distraction - I believe that a mental act cannot reach the same degree of complexity as a physical one. But even though I believe that the intentionality of mental acts is explicable in terms of (IBA), I regard it as defensible to focus on actions involving physical components in a study of the general concept of intentional action, since purely mental acts - such as visualizing, concentrating, thinking something over etc. - each raise philosophical perplexities peculiar to themselves and not germane to the general concept of intentional action.

6.3 Reason-governed Action: Complex and Non-Basic Actions.

1. The turn has now come for the definition of the intentionality of teleologically complex and non-basic actions.

I shall be brief about teleologically complex actions, because I think they can be dealt with without the introduction of any considerations new in principle. a brings about p* during t- in a teleologically complex action if: a successively makes true a set of propositions $p_1, p_2, \dots p_n$ during t- in a way such that the making true of each of them satisfies (1)-(5), and (ii) a, during t-, thinks, correctly, that the conjunction $p_1 \& p_2 \& \dots p_n$

(in the same order as they are brought about) is tantamount to p^* , at least at t_- . Intuitively, it seems reasonable to say that a teleologically basic action must be teleologically simple. (IBA) will entail this if (1) is modified to read "There is no proposition q or set of propositions $q_1, q_2, \dots q_n$ such that a , during t_- , thinks that if he causes q ($q_1, q_2, \dots q_n$) during $t_- \dots$ ". It also seems reasonable to reserve the term "teleologically simple" for what is teleologically basic, in which case (IBA) is also valid as a definition of the former notion. It should be noted that one is justified in speaking of teleological basicity or simplicity in an absolute sense, because of the reference to the agent's thought in these notions.

Let me now consider the intentionality of teleologically non-basic actions. At first sight, it might look as though this notion can be dealt with as summarily as that of an intentional teleologically complex action. It might seem that the addition of the following two conditions to (IBA) is sufficient to make it true that a intentionally brings about r in t_- in a teleologically non-basic action by bringing about p during t_- : (a) during t_- , a thinks that if he then brings about p , r will be brought about too by him, and (b) this conditional is true: by bringing about p in t_- , a brings it about that r in t_- .

But this will not do, because these two requirements are not enough to exclude what Davidson has characterized as "quaint external causal chains" ([68], p. 153). To revert to the mountaineering-anecdote: suppose that the mountaineer is correct in his belief that if he loosens his hold on the rope, he will get rid of his fellow climber. But suppose further that the result of the basic action causes the result of the non-basic action by a route different from the one the mountaineer calculated upon: he thought his companion was swinging freely on the rope, whereas it was in fact a piece of equipment that was hanging on the rope, while the companion was standing on a shelf below - something which, however, did not save him from disappearing into the abyss, because the equipment fell on him and caused him to lose his balance. Here (IBA) + (a) & (b) might well be met: the climber intentionally loosens his grip, he thinks that if he causes his hands to loosen their hold, this will cause the fall of his companion, and this thought is true. Yet it is doubtful whether he has intentionally caused his companion to fall. I am inclined to hold that his causing the fall is non-intentional, because he achieves it through a chain of lucky coincidences, and there appears to be a contrast between what is brought about by coincidence and what is intentionally brought about (e.g. meeting someone by coincidence and intentionally meeting someone). If someone (cf. [27], pp. 171-2) should be disposed to dispute this verdict and regard the act of causing the fall as intentional

(in the circumstances sketched), he must at least concede that the very fact that intuitions diverge is evidence that the case is not paradigmatic, and this is enough to justify a search for further conditions, the presence of which stamps the paradigmatic cases as paradigmatic.

The remedy that Goldman proposes is, speaking somewhat crudely, that a non-basic action is intentional only if it is "performed in the way conceived in the /agent's/ action-plan" ([88], p. 57). This takes care of the problem of wayward external causal chains in the shape it takes in the mountaineering-example, for the climber obviously did not cause the fall of his companion in the way he calculated: his action-plan did not incorporate any reference to e.g. the fall of the equipment. The trouble with this suggestion is, however, that it demands nothing by way of specificity or elaboration of the action-plan; as a consequence, it will easily become too broad, if the details of the plan are scant, and too narrow, if the details of the plan are worked out in great detail by the agent. Let me illustrate each extreme of elaboration in turn. Suppose that an absurdly self-confident golfer's plan is simply that if he swings his club in a certain way, he will hole-in-one, and that, through a stroke of luck, he succeeds in holing-in-one by swinging his club in the calculated fashion. Then, although he performed the action of holing-in-one in the manner conceived in his action-plan, he has not intentionally holed-in-one, if I am right in there being an opposition between things caused with the help of fortuitous happenings and things intentionally caused. On the other hand (cf. [202], p. 257), imagine a poisoner who thinks that if he administers a certain substance to his victims, this will cause their blood to coagulate, and hence their death, and who in fact poisons his victims to death by intentionally letting them imbibe the substance, though the substance works by blocking transmissions in neuromuscular synapses, thus causing death through suffocation. This person surely has intentionally poisoned his victims, despite the fact that he did not perform this action in the way conceived in his action-plan: he did not cause the victims' blood to coagulate.

It might look as if an impasse has been reached, that a requirement concerning the degree of detail of a true action-plan or action-recipe must founder either on the Scylla of under-specification (letting the golfer-case through) or on the Charybdis of over-specification (not letting the poisoner-case through). But the respect in which the golfer's plan is under-specified (for intentionality) is in fact different from the respect in which the poisoner's plan is over-specified (for intentionality). The poisoner's plan is over-specified in that it attempts to trace the phenomena that link

imbibing the substance with death, whereas the golfer's plan is underspecified in that he has not got sufficient background information to justify his endorsement of the conditional "If I swing the club in this way, the ball will go directly into the hole", without attaching a modifier like "possibly" to it. For the golfer is presumably not in possession of exact information concerning how the ball will be hit, how the wind is blowing, the conditions of the ground, all of which are necessary for a reliable prediction of the trajectory of the ball. Now, one might well have evidence that warrants an unqualified endorsement of a conditional, e.g. the conditional that if one lets these people imbibe this amount of that substance, they will die, without being able to explain why this is so, e.g. spelling out how the poison works. (This evidence might consist in one having repeatedly witnessed people eating this substance and dying.) In other words, one might know that the present circumstances are such that the addition of a certain fact will create a sufficient condition for a certain effect, without being capable of explaining why this condition is sufficient for the effect; one might be able to state a sufficient condition without having disclosed the mechanism of its operation.

What could be demanded, then, is that the agent believes in the truth of a set of propositions such that he is not in error in using these propositions as adequate support for the relevant action-recipe in an unqualified form. Furthermore, as the mountaineering-example shows, the backing set must be true at the action-time: this is what is wrong in this example, for the climber uses propositions, such as that his companion is hanging on the rope, to support the conditional that if he lets go of the rope then his companion will fall, and the former proposition is false.

The definition we have reached is - a little refined - this:

- (INA) a brings about r during t- in an intentional teleologically non-basic action by (intentionally) bringing about p in a basic action during t- iff:
- (1)-(5) obtain;
 - (6) During t-, a thinks that if he brings about p in t-, r will also be brought about by him in t-;
 - (7) There is a set of propositions about which a, during t-, thinks
 - (i) that it is wholly true, and (ii) that it, in conjunction with the proposition that he brings about p in t-, entails that he brings about r in t-;
 - (8) The thoughts (i) and (ii) in (7) are both true (hence, the conditional thought in (6) is also true).

Perhaps "entails" in (7) is too strong, but the relation of support must be firm enough for it to be legitimate to compress (6)-(8) into the statement that a knows that the conditional in (6) is true. For, harking back to the discussion in 4.2.3, it must be proper to say that a successfully puts his knowledge how to achieve an effect into practice when he intentionally achieves it in a non-basic action. When one brings about something intentionally, one does not rely on luck, one exercises a skill.

2. Are conditions (1)-(8) really sufficient or is it necessary to add (9) During t-, a wants_d to bring about r in t-²³ to them?²⁴ There is a distinction between what is done intentionally, on the one hand, and what is done (merely) knowingly or wittingly, on the other, and, it seems, a brings about r only knowingly or wittingly if (1)-(8), but not (9), are true. It might be objected, however, that (9) is really a superfluous addition, since it is inferable from (2) and (6). For I have affirmed the Kantian principle that whoever wants_d an end must want_d whatever means he considers necessary to this end, and if p is sufficient for r, r is necessary for p; therefore, since a wants_d p, he must want_d r as well.

The conclusion of this argument is, however, counter-intuitive: it seems odd that if I want to wear a certain pullover, I must necessarily want to wear it out (to some extent), if I recognize that the former will inevitably lead to the latter. Would one of my desires be frustrated, if, per impossibile, I succeeded in wearing it without wearing it out the slightest bit? Clearly not. Of course, if I voluntarily wear the pullover, then, having the knowledge I have, I cannot have an aversion to wearing the pullover out that is stronger than my desire to be clothed in it. I must accept the wear and tear as a price for wearing it, but I need not positively want_d it: after having worn it, I need not be disposed anxiously to look for some damage in the fibres and, when finding some, to sigh with relief.

The argument for the redundancy of (9) is, then, fallacious, but where more precisely is the fallacy to be located? The Kantian principle might seem the most suspect premise, but I hope to be able to demonstrate that the falsity resides in the assumption that if bringing about r is a necessary condition for bringing about p, then it is a necessary means for bringing about p. Wearing a pullover out (to some extent) is a necessary condition for wearing it, but it is not a necessary means for wearing it. To put the lesson generally: the relation of being a means to an end cannot be reduced to that of being a condition for a state of affairs, because the latter is reversible in a way that the former is not: if p is a sufficient condition for r, r is a necessary condition for p, but if (causing) p is a sufficient

means for (causing) r, then it is seldom - I am inclined to say never - true that (causing) r is a necessary means for (causing) p.

What is the ground of this irreversibility of the means-end relation? The first hypothesis that presents itself is perhaps that the irreversibility is rooted in the asymmetry of the causal relation: the means - or the result of the means-action - must cause the end. But this cannot be right, for there are cases when the end causes the means-result: e.g. when one achieves the end of causing a neuron-firing down an efferent pathway by means of lifting one's arm.²⁵ Admittedly, there are philosophers - e.g. von Wright ([217], pp. 76-81) - who hold that the arm's going up here causes the neuron-firings and that we here, consequently, have an instance of backwards causation, but this seems a paradoxical interpretation. (By the way, this case vividly illustrates the need for a distinction between teleological and ontological basicity, for while causing neuron-firings are ontologically basic relative to causing the arm to rise, causing the arm to rise is teleologically basic in relation to causing neuron-firings.)

Instead of being rooted in causal asymmetry, the irreversibility, I suggest, has its roots in what one might call considerations of epistemic accessibility: a, at t, can consider causing p to be a means of causing r only if he at t expects, as he has been used to, to be able to verify that he has caused p prior to his verifying that he has caused r. This implies that it is only because one has no concurrent awareness of neuron-firings that one could raise one's arm in order to cause them; if neuron-firings had a more immediate impact on one's consciousness than arm-risings, one would rather cause neuron-firings in order to raise one's arm. To my mind, this does not seem implausible.

In the light of these remarks, a reformulation of (1) and (6) seems called for: it should be added that, for a, q must be epistemically more accessible than p and that p must be epistemically more accessible than r. However, instead of explicitly appending these clauses, I adopt this convention regarding the formulation of conditionals of the form "If a brings about ..., ___ is brought about too": they should be so formulated that the epistemically more accessible proposition appears in the antecedent, unless otherwise explicitly indicated. Nothing hinges on whether this is taken as a stipulation or description of ordinary language.

Granted this convention, (9) is by no means superfluous, for if a is able to establish that he has brought about p - i.e. satisfied his desire - before establishing that he has brought about r, he cannot want r to occur as a sign that he is in the process of gratifying his want for p. Instead, being aware

that he has satisfied his desire, he might view the question whether r follows with indifference.²⁶ This amounts to saying that my pronouncement, that if a thinks that if he causes p then r will be caused too, and he wants r, he has a reason for wanting and causing p, can only be accurate if the conditional that expresses the thought-content is formulated in compliance with the convention mentioned. My wanting to exercise and my conviction (unorthodoxly formulated) that, if I cause myself to sweat then I am exercising, give me no reason for wanting to cause myself to sweat, since I do not depend upon discovering that I cause myself to sweat for determining whether I exercise.²⁷

If we suppose that (6) is a's reason for wanting and causing p during t - and hence that (9) is satisfied as well - we have in (INA) a definition of a reason-governed action, i.e. an action performed because of and in accordance with a certain reason, in such a way that both the proposition constituting the protasis and the one constituting the apodosis are intentionally made true by a basic action and its effects.

(INA) also offers material for a definition of a more restricted concept of action than the one examined in chap. 2, a concept in the vein of authors like Chisholm ([52], pp. 54, 64), Davidson ([66], pp. 6-8) and Goldman ([88], pp. 17-8, 44-8, 63-72). An action could be an intentional basic action together with whatever is brought about by means of it, whether intentionally or not. Or one need not even - like the aforesaid philosophers - require the presence of an intentional basic action: an action could be just an attempt, whether successful or not, at a basic action and whatever is non-basically made true by means of this attempt. In this case, nothing need be intentionally caused for an action to have taken place, though something must then be unintentional (in contrast to the action-concept in chap. 2). Every action (in the broad sense) which qualifies as an action in at least the latter and weaker of the senses just outlined, will be referred to as a teleological action.

6.4 The Domain of Application of (IBA).

Which action-types are those individual actions which are denoted by the term "(teleologically) basic action" instances of? Or, in other words, which types are those individual actions that fall within the applicability of (IBA) exemplifications of? As will transpire, this question is by no means easily answered. Answering it will, however, bring along an additional advantage (beside clarifying what concrete cases to test the definition by), viz. that of removing the ground for the complaint that my account of the thought-guidance of intentional action imputes too elaborate and detailed

an awareness to the agent of what is caused, for it to be realistic. Surely, it might be objected ([199], pp. 251-2), when one types this page, one does not think of every finger-movement one performs in the process of depressing the keys. Such an awareness would distract one's attention from the task of expressing one's thought and would make one's typing extremely slow.

Bodily actions, moving parts of one's own body, are certainly among the things that can be done in a basic way, though one can move parts of one's body in a non-basic way as well, e.g. if one moves one arm by pushing it with the other. But in order to reveal what is distinctive about moving one's own body, it will prove necessary to go one step further and assert this: in the ontogenetic development of human beings, moving one's body as a basic action is prior to moving it as a non-basic action. It is not the case that one acquires the capacity e.g. to move a finger as a basic action by engaging in training consisting in the putting into practice of a recipe of the form "If I bring about ..., my finger will move" or, in other words, that one gradually becomes capable of moving a finger as a basic action by repeatedly moving it as a non-basic action. The argument for this contention about bodily actions is this. There must, for every agent, be action-types instantiations of which in basic action take precedence over instantiations in non-basic action; this is a corollary of the fact that every non-basic action presupposes a basic action which generates it. And surely no action-types are more fit for being of this kind than bodily action ones, for no object in the material world is more intimately related to a person than his body.

It should be emphasized that claiming that, for a certain agent and for a certain type of actions, basic exemplifications have priority, is not incompatible with the agent's having acquired the capacity basically to execute actions of this type by practice or training. For there is a distinction to be drawn between what have been called "performance-recipes" and "acquisition-recipes".²⁸ A performance-recipe is what I hitherto have called simply a recipe, i.e. a conditional informing the agent of means available for him for realization of a certain state of affairs. Acquisition-recipes are really a subclass of performance-recipes: they inform the agent of the means for achieving a certain type of state of affairs, namely the type consisting in the agent possessing a certain ability to act. Note that since an acquisition-recipe does not tell one about the means for achieving the result of the action the occurrence of which constitutes the exercise of the ability, the fact that an agent has acquired a certain ability by practice, consisting in the application of an acquisition-recipe, does not entail that the action-type, the exemplifications of which constitute exercises of this ability,

was first exemplified in non-basic actions. Coming into possession of a capacity by training, consisting in the application of an acquisition-recipe, is quite consistent with the capacity right from the start (i.e. when it first began to be exercised in teleological action) being exercised in basic actions. As an illustration, think of someone practising in order to be able to reach his toes with the tip of his fingers without bending his knees. His acquisition-recipe will presumably run along these lines "If I keep on bending my torso forward as far as I can without bending my knees, I will finally be able to reach all the way to my toes". Suppose that this is correct and that the agent one day finds his finger-tips on a level with his toes; then this will still be a basic action, even though it has been preceded by training. Bending forward as far as he could would on this occasion simply be reaching all the way to his toes, it would not be a successful means to it.

Only, it has to be remembered that not every capacity whose exercises are originally basic can be acquired by training. For training or practising presupposes that one already possesses the ability to perform some teleological actions (and hence some basic actions): training is doing something with the purpose of acquiring (or maintaining) an ability. (The training designed to establish a capacity of the basic kind has, I think, two major forms. One is already outlined in the text: when the training has the purpose of increasing the strength of one's muscles or the agility of one's joints. In the other form, one is already able to do the basic action, though only as a part of a more complex action (e.g. wiggling one's ears as a part of contorting one's face), and training takes the form of repeatedly performing the complex action, while excluding more and more components.)

The upshot of the reasoning so far is this: for every agent, there are types of action - among which bodily action types figure - such that instantiations of these types in basic actions of the agent are genetically independent of instantiations of the same types in non-basic actions of the agent. In other words, the agent has not learnt to do them basically by doing them non-basically, although he has perhaps executed other non-basic actions in order to gain possession of some of these abilities to act basically.

Now if an action of a's involves the making true of an antecedent of a recipe of a's such that the antecedent is constituted by an action-type which is originally exemplified by a in basic actions, then (IBA) must yield a sound verdict concerning the intentionality of the present instance. So, the domain of applicability of (IBA) covers at least bodily actions basically performed.

However, I regard it as an incontestable fact that the repertoire of basic actions of a normal mature person ranges far beyond his bodily actions (cf.

[134] & [171]). Circumstantial actions can be executed as basic: one can put a finger on a key of a type-writer without being aware of the finger-movement in itself, e.g. which finger one uses; one is aware of what one is doing simply as putting a finger on a certain key. Manipulative actions can also be performed as basic: one is aware of what one is doing simply as depressing a certain key. The same holds even for manipulative actions which are consequential: in typing one need not be conscious of what one is doing otherwise than as typing a certain letter on a sheet of paper. In this way, a basic power can extend itself further and further beyond the component of bodily motion. But alongside this "vertical" extension, there is also a "horizontal" extension, a widening of a basic power to encompass actions of greater and greater ontological complexity: e.g. an experienced typist is not conscious of typing individual letters, but only complete words or even larger linguistic units.

Recognition of these processes of deepening and widening one's repertory of basic actions should undermine the charge that my account of the thought-guidance of action suffers from an unrealistic proliferation of episodes of thought or awareness. For this objection can, I believe, only be sustained so long as it is nurtured by the false assumption that my doctrine is committed to the thesis that the agent's awareness is always, to some extent at least, riveted to what is relatively basic and simple in the ontological sense, e.g. bodily movements. Such a riveting would indeed make impossible e.g. the rapid typing of an experienced typist. But I am by no means forced to deny that, with learning and experience, the focus of consciousness tends to move "outwards" and to encompass larger sequences of results so that no separate attention is paid to the results of simpler or more basic actions, which were formerly to the fore of attention. I must point out, however, that this, according to my analysis of intentional action, implies that actions which were once intentional become non-intentional, habitual.²⁹

Unfortunately, there is a price to be paid for this enhancement of the plausibility of my theory of intentional action: the selfsame fact of the extension of a basic repertoire creates complications for (IBA). Suppose that a thinks of what he is causing, not as a certain finger moving to a certain key and depressing it, thereby causing an imprint of "a" to appear on the paper, but simply as an imprinting of "a" on the paper. Suppose further that a succeeds in causing the imprint of "a" to appear; he has then caused the imprint to appear in a (teleologically) basic action. The action, therefore, falls within the province of (IBA). Now it is quite obvious that the motivational background of the action could easily be filled out in a way that

would satisfy (IBA)'s conditions for a's causing of the imprint to be intentional, but in a way which would not preclude circumstances which render the action non-intentional according to the verdict of intuition. Imagine e.g. that a in fact depresses the key marked "s", but that he still succeeds in typing an "a", because a mechanic, in repairing the type-writer, has, unbeknownst to a, committed the mistake of attaching the "a"-type to the "s"-key. I think it clear that circumstances like these, the absence of which is obviously not guaranteed by (IBA), will make one inclined to refuse to call the action intentional. In fact, what confronts us here is essentially the problem of wayward external chains all over again: not surprisingly, this complication begins to infect the execution of basic actions, when these, so to speak, conquer territory formerly belonging exclusively to non-basic actions.

Fortuitous happenings can infiltrate even basically executed circumstantial actions: suppose a drunkard thinks, correctly as it happens, that what he is doing is putting his right index-finger on an elevator-button, but that his finger would have missed the button, were it not for the fact that the elevator, in which he is standing, had made an unexpected slight jerk, at a crucial moment. Again, it is reasonable to say that the drunkard's basic action was not intentional, although (IBA) might well be satisfied.

A way of modifying (IBA) so as to avoid these counter-instances has been adumbrated above: to restrict the validity of (IBA) to the intentionality of exemplifications of action-types whose instantiations in basic actions for the agent are genetically independent of instantiations in non-basic actions. The types "typing 'a'" and "putting a finger on a button" fall under the categories of manipulative and circumstantial actions, respectively, and, it might be urged, instantiations of types from these categories in a basic manner are generally dependent on instantiations in some non-basic manner.

In the case of manipulative actions, especially if they are the consequential kind, this thesis seems hard to dispute: it appears obviously true e.g. that someone, in the first phases of the process of learning to type, types letters by means of consciously and carefully directing his fingers to different keys and depressing them and that an ability to type letters, basically, develops as these finger-movements become habitual. But the thesis appears easier to challenge in the case of circumstantial action: is there really a phase in the development of every individual when he e.g. moves his arms, hands, fingers in order that the latter should make contact with a button, where these bodily movements in themselves are the objects of intense concentration, detailed awareness? It might be said that an infant only moves

his limbs when he seeks contact with some perceived object. But that is beside the point, for the issue concerns what awareness he then has of his bodily movements. If putting a finger on a button in optimal circumstances, for someone who has not acquired the knack of doing it, is comparable to putting a finger on a button in tricky circumstances (when one is intoxicated, the button moves swiftly and unpredictably etc.), for someone who has acquired the knack of doing this in optimal circumstances, then I am inclined to return a "yes" to the question. I am also persuaded by the comparison.

If this is correct, (IBA) can be saved by restricting its domain of application in the indicated manner, for, so far as I can see, originally basic actions are immune to the malaise of extravagant external causal chains. Of course, if this strategy is chosen, (IBA) will have to be complemented with an account of the intentionality of basic actions falling outside its field of application. But the essential material for such an account should already be contained in my discussion of wayward external chains in connection with non-basic action (see 6.3.1) - as noted above, the difficulties are caused by the reappearance of this problem in the realm of basic action. a's causing p in such a basic action takes place at will in t- iff, in addition to the conditions of (IBA), a during t- has (or once in his history had) knowledge of the truth of a conditional the consequent of which is constituted by p, and the antecedent of which is constituted by the description of an action-type which (i) a instantiates during t-, though mechanically (hence, a does not, during t-, call the conditional to mind and apply it), and which (ii) a can instantiate in a basic action the intentionality of which falls within the jurisdiction of (IBA). According to this, the typist who luckily types an "a" in a basic action, has not typed it at will, since he, being ignorant of the mechanic's mistake, does not know that finger-movements of the type he causes will in the present situation result in an "a" being imprinted on the paper. For similar reasons, the drunkard puts his finger on the button in a non-intentional basic action: he cannot know that a finger-movement of the type he brings about will lead to the finger coming into contact with the button, since he does not know about the jerk, without which the finger would miss its target.

However, should anyone insist that the restriction I have imposed on (IBA) - confining it to originally basic actions - is insufficient to exclude e.g. circumstantial actions because they, too, are performed as basic from the very moment when they are first performed as teleological actions, I still have means for saving (IBA) at my disposal. I could strengthen (3) of (IBA) from that a thinks that he might (at least) be able to cause p and thinks that

what he is causing will possibly be p, to that he knows these things, knows that he has the ability and knows that what he is causing will be p. (Knowledge here again implies the existence of adequate backing that is true.) But I am reluctant to resort to this measure, because it seems to me to have the drawback of demanding too much in the case of bodily action. A person might not know whether he can e.g. wiggle his ears or move his hair back and forth in basic action, but might merely try in order to find out; still, if his attempt is successful, it appears reasonable to me to say that it was done at will, despite the lack of knowledge.³⁰

Therefore, I am personally more sympathetic towards the first strategy for rescuing (IBA). But I do not feel ready to take a definite stand on the matter, because I am by no means sure that I have fully gauged the intricacies of the basic/non-basic action distinction. It is tempting, in the face of these complexities, to share Annette Baier's [21] skepticism about the feasibility of the dichotomy, but I think it should be maintained, in spite of its involved and problematic nature, because it seems to shed much light on the experience of acting.

CONCLUSIONS AND LIMITATIONS

The main conclusions of this work can be set forth fairly briefly. I have striven to establish that, though what is a reason for action for an agent is dependent on his wants, reports or expressions of these wants do not constitute part of the content of these reasons. Instead a's reason for bringing it about that p is, when phenomenal, constituted entirely by a content of his thought expressible in the form "If (only if) I bring about p, q is brought about by me as well", where q, however, must satisfy the requirement of being wanted by a. According to the analysis of conditionals in 4.1, a's thinking this thought presupposes that he regards it as an epistemic possibility that he will bring about p - and, a fortiori, that he regards it as an epistemic possibility that he is able to bring about p. Now this reason governs his bringing it about that p - assuming this to be a basic act in the sense explored in 6.4 - just in case his thinking that it is epistemically possible that he has this ability to cause p causes - because of the conditional connection thought to obtain between p and q - something to become a fact which a correctly takes to be p, provided that this taking on a's part plays an essential role in the causing. In terms of the analysis of wanting set forth in chap. 5, this is sufficient for saying that a wants_d to cause p - hence, that his attitude of wanting has been transferred from the consequent of the conditional to its antecedent. On the account of wanting_d something only indirectly obtainable, this transferral - assuming that the means are necessary as well - takes place by logical necessity. Finally, a's causing of q in a teleologically non-basic action, via his causing of p, is governed by the aforesaid reason just in case these conditions hold in addition to those mentioned: the reason-constituting conditional is true, and, among the convictions a uses to support this conditional, there is a set of true propositions that is sufficient for the unguarded assertion of this conditional - or, to express it in condensed form, a knows that the conditional is true.

If this theory is acceptable, an account has been given of what it is to have a reason for action in one's mind and of what goes on when this reason is translated into - basic and non-basic - action, in a way that renders the latter two intentional, an account that at no point appeals to sui generis action-theoretical concepts, such as immanent causality and volitions (see (1) below). In my analysis, the notions of thinking true or false thoughts and of causation figure at the fundamental level; but, although a full explication of these concepts has not been possible here, I have tried to show that they do not rest on action-theoretical notions. Thus, I regard myself as

having succeeded in my stated aim of presenting an elucidation of that segment of the practical sphere which is constituted by the concepts of having a phenomenal reason, of moving from ends to means in reasoning and of acting intentionally, which does not involve any reference to irreducible or unanalyzable action-theoretical concepts. It seems, then, that action-theory can be carried out without adding concepts peculiar to it to the vocabulary of primitives.

The real focal point of the thesis is the analysis of reason-governed actions; the study of the notion of a phenomenal reason is strictly only a prolegomena to this. For the test of an analysis of the latter sort lies in the plausibility of the account it allows of the notion of acting for a reason (and in - what I have not been able to enlarge upon here - its extension to real reasons). My focal question has justifiably been called - by Myles Brand [37] - the fundamental question of action theory. Brand formulates the question thus: "What properties must a mental event have in order for it to be the proximate cause of action, or more generally, the proximate antecedent of action?" ([37], p. 139). This formulation is not felicitous. Its most important defect is that two distinct problems are rammed together in it, namely the problems of specifying what must obtain for an action to have occurred and what must obtain for an intentional action to have occurred. In order for an action to have taken place, it is not necessary that any mental event at all occurs: a has acted if a fact of change - be it mental or physical - pertains to a, and this causes a state of affairs, say p, to obtain. But a's causing of p amounts to an intentional (basic) action only if the causing facts pertaining to a are of the mental type, viz. only if they include (at least) a's thinking that he might have the capacity to cause p and that what he causes is the realization of a state of affairs - namely p - that constitutes the result of an action exercising the capacity he has tentatively ascribed to himself. The feature of these mental facts, thoughts, which turns the causal sequence they constitute together with their effect into an intentional (basic) action, is that there is a correspondence between the content of the latter thought and its effect, i.e. the content refers to and is true of the effect. This is my answer to Brand's question, given that it concerns intentional actions and the crucial case of basic ones.

I hope that this theory of having reasons in one's mind and acting on them has been expounded in sufficient detail for it to appear intelligible. But considerations of space have imposed certain limitations: besides perhaps forcing me to be a little briefer on some points than I would have liked, they have necessitated the omission of treatment of topics which would have under-

pinned and illuminated the theory further. Let me end by listing the most urgent of these topics.

(1) My reductive analysis of the concept of a reason-governed action ought to be supplemented with a critical examination of non-reductive analyses, though the burden of proof - i.e. the need to show the indispensability of some irreducible practical or action-theoretical notion - is on the non-reductive analyst. To some extent, I have fulfilled this obligation by my brief criticism of the notion of immanent causation in 2.2. But there is another type of non-reductive analysis that is much more popular today - and hence much more in need of being argued against -, viz. the volitional type of theory. By a volition theory I mean a theory that includes these two doctrines: (i) there are mental episodes of a distinctive action-theoretical or practical kind - e.g. thoughts of a practical content, tryings, acts of will - and (ii) these episodes are determinants of (results of) teleological actions. This type of theory is presently espoused in many different forms by many different philosophers: by Sellars (e.g. [180], pp. 150-9; [181]; [184], pp. 189-95; [186], pp. 233-6; [188]) and philosophers influenced by him like Aune ([13], chap. 2), Castañeda ([45], chaps. 6 & 10), Tuomela ([205], pp. 136-43), and by other philosophers like Brand [37], Davis ([72], pp. 15-26), Goldman ([88], chap. 4; [89]), Kenny ([111], chaps. 10 & 11; [112], chap. 3), McCann [125], O'Shaughnessy [150] and Whiteley ([212], pp. 48ff). Since the volition theory takes many forms, it is more or less impossible to find an argument that hits them all - except that of the challenge presented by the elaboration of a reductive account. Therefore, I will here focus on a particular version of it, to wit the one propounded by Sellars and his school.

According to Sellars (e.g. [180], pp. 150-9, [181]), an intention is either a disposition to think thoughts of the form "I shall do A at t" or an actual instance of such a thought; decisions are the subclass of such thoughts that occur as the culmination of deliberation ([180], p. 156), and volitions are the subclass where "t" is "now". Aune prefers ([13], pp. 63ff) the auxiliary verb "will" to "shall", and Castañeda discerns a common deep-structure underlying both declarations of intentions (decisions and volitions) and imperatives, a structure expressible by the infinitival construction "X to do A at t" ([44], pp. 144-6; [45], pp. 169-72), but these divergences need not detain us now.

There are two critical comments I would like to make about this version of the volition theory. The first concerns whether an appeal to the causal activity of a volition in the genesis of action can block Davidson's wayward

causal chains that disrupt intentionality. Proponents of the model seem to assume so (see e.g. [13], pp. 101-2; [45], pp. 282-3; [184], pp. 193-4; [205], pp. 257-8). Yet, it appears to me, it has never been satisfactorily spelt out how the presence of a volitional thought guarantees intentionality of the appropriate behaviour caused.

So far as I can find, the only attempt to dissect the distinctive causal properties of the volitional thought is that of Castañeda ([45], chap. 10, sec. 3). Castañeda is fully aware (p. 282) that contemplation of any proposition may accidentally set off a causal mechanism that makes the proposition true. If the causation is to amount to an intentional act, he appears to reason, there must be something in the thought-content that marks that the content is presented as something whose materialization is at stake. To mark this is, if I understand Castañeda, the function of the volitional copula: "a practical copula is ... the ordinary propositional copula with a signal of its practical or causal openness" (p. 280), or "(i)t is a focussing mechanism that spots the relevant parts of a proposition where the agent's thinking causality can connect the proposition whose truth he is attempting to bring about with the mechanisms that carry out the attempt" (p. 281). Undoubtedly, there is a considerable resemblance between this and my requirement that what is intentionally done must be presented to the agent under the modality of something he can bring about. Now why not simply rest content with fully exploiting the disclosed modal aspect in a doctrine such that the agent non-coincidentally thinks that what he does is exercising the ability? It seems to me that when Castañeda goes further and attempts to justify the introduction of an elusive practical copula, he produces only metaphor and rhetoric (e.g. that the copula is the "opening" where causality enters, that the causality is "internal", p. 283).¹ Thus, I cannot but conclude that Castañeda's theory has not been developed into a serious rival to my reductive account.

The second critical point cuts even deeper because it argues not merely the superfluity of volitions in the intentional causation of appropriate results, but also in the causation simpliciter of such results. More precisely, it denies that volitional thoughts are a motivational force independent of attitudes of wanting. How does it come about, according to the volition theory under scrutiny, that a volition to bring about p - a thought expressible by "I shall bring about p now" - is regularly followed by the bringing about of p? Aune gives this answer (cf. [180], p. 158; [184], pp. 192-5):

"Part of what is involved in learning to think in a volitional way

is that we develop a propensity to make appropriate movements in response to thoughts like 'I will now do such and such' ([13], p. 64).

(The other part involved in learning to use the volitional formula is to master the use of it in practical reasoning.) But what motivates us to develop this propensity in response to an occurrence of something of the form "I shall now bring about p" in one's mind? Surely, the mere fact that something that could be rendered by "I shall now bring about p" occurs to one is not in itself enough to motivate one to bring about p. To assume that how this propensity to bring about p arises is unproblematical is, in effect, to assume what is to be accounted for: if it is obscure how a volition can cause behaviour in conformity with it, it is also obscure how it can lead to a development of a propensity to such behaviour.

A more plausible explanation of how the link between the volitional thought and the (propensity to) appropriate behaviour is established takes its departure from the assumption that the propensity is present in the pupil prior to the introduction of the verbal phrase, rather than that the propensity is created by this introduction. Language-learners are instructed to use "I shall now bring about p" when and only when they already have sufficient motivation to bring about p, i.e. they want to bring about p strongly enough to bring about - or at least try to bring about - p. They are trained to correlate this verbal form with an independent motivational state which they are aware of and which is by itself sufficient for action. On this view of the matter, the thought that one shall make it true that p is superfluous from the motivational point of view; it is conditioned to occur when and because one is already set on achieving that p. It is (roughly) an expression of the motivational state that "I want to bring about p strongly enough to (try to) bring about p" is a report of. The expression contributes as little to the causation of behaviour as does the report. They are both reflections or side-effects of a motivational state, a dominating attitude of wanting, which culminates in action. Compare this with the state of feeling pain: the expression "Ouch" and the report "I feel pain" are not constituents of the state of feeling pain. One uses them - and was instructed so to use them - when and because one already, beforehand, feels pain. And they are not parts of what causes pain-behaviour.²

This account, if sound, disposes of (ii) above: volitions - at least if construed as shall-thoughts - are not (causal) determinants of (the results of) actions. Since the term "volition" traditionally covers the conjunction of (i) and (ii) it had better be discarded. Still, (i) might be true: nothing has been said to disprove the existence of mental episodes of a distinctive

practical kind, e.g. in the shape of shall-thoughts.

I am prepared to countenance the reality of thoughts the content of which is properly couched in "I shall bring about p at t", because I think Sellars is correct in maintaining that we need access to them in the explication of the notions of intention and decision. An approximative definition of the concepts of intention and decision as I conceive them runs like this:

(DI) a intends to bring about p =df

a is in a state which at its onset was adequately expressed by his thinking a thought the content of which is expressible by "I shall bring about p" (i.e. a decided to bring about p) and which is still correctly so expressed.

The state referred to in (DI) is roughly the state of wanting to bring about p so strongly that one actually brings about - or at least (if one lacks the ability) tries to bring about - p. I think this is inaccurate only in one respect: while wanting to bring about p only requires that the agent takes it to be an epistemic possibility that he can bring about p, it is arguable that intending/deciding requires one to be more or less certain of one's ability. If one is less than certain, it is perhaps more correct to say that one intends to try.

(DI) is something of a compromise between two camps. In conformity with Sellars' approach, it includes mention of shall-thoughts, but contrary to Sellars, it also insists that these thoughts must be expressions of an independent state of (roughly) wanting_d. This latter feature brings my view close to that espoused by Audi ([9], pp. 389-96), Beardsley ([27], pp. 176-80) and Churchland ([55], pp. 230-1), according to which an intention to bring about p is (roughly) reducible to a dominating want to bring about p and a belief that one will bring about p, although it is still distinguishable from these writers' view by the reference to shall-thoughts. I am inclined to regard this reference to linguistic expressions, shall-constructions, as essential because of my intuition that, while it is perfectly in order to ascribe wants and beliefs to non-language users like animals, it sounds odd to attribute intentions and decisions to them.³

According to (DI), the concepts of intention and decision occupy an intermediate position between a motivational concept like that of wanting and a communicative one like that of a declaration or an avowal. There is a link to motivation in the connection to wanting, and there is a link to communication, to linguistically declaring or expressing (not stating or reporting) states of mind, in the reference to shall-thoughts.⁴ A creature cannot logically intend or decide unless it has acquired the capacity to

express its motivational states in shall-sentences, though of course it need not announce every intention or decision overtly: it is enough if the verbal conditioning manifests itself internally, in thoughts. Note that this view of intending and deciding, though it acknowledges (i), i.e. thoughts of a distinctively practical kind, does not compromise my reductive account of acting intentionally or voluntarily,⁵ since the practical thoughts are motivationally impotent, mere reflections of motivational states - of wanting - that by themselves culminate in action. The practical modality of thought belongs entirely to the dimension of communication.

This view of intention dissolves a puzzle that threatens the alternative view that intentions are motivationally efficacious. It is widely held that "a intends to bring about p" entails "a believes that he will bring about p" (where "will" carries predictive force).⁶ On my view the explanation of why this entailment holds is as follows. As remarked, "a intends to bring about p" entails "a believes that he has the capacity to bring about p". From the former one is also allowed to infer "a believes he wants p sufficiently to try to bring it about", since intending includes giving expression in thought to - and hence being aware of - such a state of wanting. But if one believes both that one has the ability to cause p and that one is sufficiently motivated to make use of this ability, then one must believe that one will bring about p.

Suppose, on the contrary, that intentions are motivationally potent forces over and above wants, suppose, more precisely, that intending to cause p on one occasion is a sine qua non for causing p. Then believing that one will bring about p on that occasion requires one to believe that one intends to bring about p on that occasion. But then, given the entailment endorsed in the foregoing paragraph, a cannot intend to cause p on that occasion unless he believes that he intends to cause p on that occasion.⁷ But this seems to make impossible the formation of the intention, for normally one can hardly believe one intends p without already intending p, and, ex hypothesi, one cannot intend p without believing that one intends p. The difficulty deepens if we try to make room for the possibility that a gains insight into the logical analysis of intending: the content of his belief is then that he intends p, i.e. (among other things) that he believes that he intends p, i.e. that he believes that he intends p ... and so on ad infinitum. A belief with such a content can scarcely be held. These perplexities are altogether avoided, if intentions are not conceived as separate motivational thrusts, since believing that one will bring about p will then not require any beliefs about one's intentions.

(2) An adequate theory of reasons for action must give a plausible account of such reasons not only in explanatory contexts, but also in evaluative ones. In my concept of a real reason I have, I think, laid the foundations for an extension of my theory into the evaluative domain, but I have not worked out and defended this extension in detail. Someone might regard this omission as quite symptomatic, thinking that an account which makes reasons dependent on wants, though perfectly at ease in explanatory contexts, comes up against difficulties in evaluative ones. I can do nothing to counter this general presumption here. What I can do is to indicate, by deflecting a more specific argument, how an appeal to real reasons can be used to fortify my analysis of phenomenal reasons.

According to Kurt Baier an account which lets reasons for action rest on desires has the consequence that it becomes impossible to justify or criticize (in a word: to evaluate) a person's "ends and aims" in terms of reasons, if these ends and aims are ultimate, i.e. not means to further ends and aims ([24], pp. 87-8, 91).

What could be meant by "ends and aims" here? A plausible interpretation seems to be that a person's (ultimate) ends and aims are constituted by the states of affairs he wants for their own sakes, i.e. independently of any phenomenal reasons. But then it is not true that a theory of reasons like mine excludes the possibility of evaluation of ends in terms of reasons, for though a person, in his deliberations, must accept something as an end, i.e. must want something solely for its own sake, without considering its consequences, that is, must want it independently of phenomenal reasons, nothing prevents there being consequences of this end that are relevant to his wants, i.e. real reasons for or against wanting the end. Normally, there will indeed be such consequences, and occasionally they might even be of such weight that if the person took them into account, his overall attitude towards the end would no longer be one of wanting_d. In short, a person - who avoids perishing from inactivity - must accept some states of affairs as ends in the sense of something he wants quite apart from phenomenal reasons. But it is gratuitous to assume that these (or any) states of affairs are also ends in the stronger sense that it is inconceivable that he should want them also partly because of reasons, were he to explore their consequences, i.e. that the chain of real reasons, just as the chain of phenomenal reasons, must have an end-point, and at that, that these end-points coincide.

These points are perhaps sufficient to ward off the suspicion that the doctrine that reasons are dependent on wants inevitably leads up to a too simplistic view of reasons in evaluative contexts. Most people are, I believe,

prone to think that deliberation and justification of actions are complicated and perhaps even ultimately inconclusive matters - something reaching far beyond merely establishing what are the means for a given end. But the seeming endlessness of the chain of real reasons is precisely what is needed, it appears to me, to cater for this complexity and inconclusiveness.

Still, there can be no doubt that some philosophers (see e.g. [170], pp. 75-9) would want to maintain that some reasons (e.g. moral ones) are objectively binding on people, i.e. binding irrespective of their desires. These philosophers would claim that a man, however callous and insensitive he might be to the needs of others, nevertheless has reasons for acting altruistically or benevolently. I am not trying to deny the legitimacy of such a conception or even to deprive these philosophers of the phrase "reason for action" to mark it (though personally I think that e.g. "obligation" is more suitable); I am merely insisting that they must acknowledge reasons in my sense as well - otherwise the arguments raised against Baier in 5.1 are actualized again.

(3) A second difficulty Baier raises for a view of reasons like mine seems even more threatening. Baier claims that it is "unable to account for the cases in which someone, in the full knowledge of the facts and the full knowledge that his behavior would be contrary to reason, nevertheless acts in this way" ([24], p. 142). In other words, my theory of reasons implies that if a intentionally brings about p at t, then, at t, he had stronger phenomenal reasons for bringing it about that p than for anything incompatible. It is impossible that he, if he at t were aware of his phenomenal reasons as reasons operating in his mind, could truthfully say "According to the facts I am aware of, I have the best reasons for causing p now" and still, simultaneously, intentionally cause something incompatible with p. But, it may be urged, this is precisely what happens in the common case of akrasia or weakness of will. Hence, my theory of reasons is refuted.

No doubt, it is tempting to respond to this consideration by loosening the tie between reasons and wants so that it becomes possible simultaneously to judge, correctly, that one has best reasons for doing one thing and still to want_d another.⁸ But the arguments put forward against Baier in 5.1 and against the normative conception of rules of reasoning in 6.1.1 militate against taking this way out. My reply will instead be that the objection misconstrues the notion of weakness of will and that, if the notion is properly understood, it will be seen to be in harmony with my theory. Before I start unfolding my doctrine of weakness of will, let me make it clear that I am here, in order not to introduce any irrelevancies, assuming something the truth of which I am not entirely certain, namely that e.g. Davidson ([65], p. 94) is correct

in that all actions resulting from weakness of will are intentional.

The phenomenon of weakness of will occurs primarily, I suggest, when the subject is affectively disturbed, when he is subject to certain sensations, such as those of pleasure or pain, or emotions like fear, compassion etc. Let me here concentrate on the comparatively simple phenomenon of a sensation of physical pain. Such a sensation has obvious effects on external behaviour: it tends to produce shrinking, grimacing, crying etc. But it has also effects on thought: it tends to rivet the subject's attention to the sensation. (This effect on thought is, I argued on pp. 107-8, in more advanced organisms necessary for influence on behaviour.) As a consequence, states of affairs which were formerly the object of the subject's thought might no longer be so.

These considerations provide the material for an interpretation of weakness of will that is consistent with my theory of reasons for action. a's presently wanting_d p is an outcome of weakness of will just in case his presently having stronger phenomenal reasons for causing p than for causing anything incompatible and, hence, his wanting_d p, has succeeded - due to his having entered into an affective state which blocks his thinking of certain states of affairs he previously thought of - his earlier state of having stronger phenomenal reasons for something incompatible with p. a acts out of weakness of will just in case this want_d of his that p determines his causing of p in such a way as to make it intentional. It is evident that this account of weakness of will is harmonious with my theory of reasons, because it does not imply that the akrates, the weak-willed person, acts contrary to the thrust of the phenomenal reasons he has at the time of action, but merely to the thrust of the more inclusive phenomenal reasons he had at a prior time and which he will presumably have again at a later time when he is no longer in the grip of feelings. On this conception, the weakness of a man's will consists in the liability of his phenomenal reasons, and hence his attitudes of wanting, to change - more precisely, to become less inclusive in scope - as the result of affective pressures, whereas the strength of a strong will consists in a relative constancy of phenomenal reasons through varying affective conditions.

(4) This brings up the notion of an emotion again. There are several reasons - beside the relevancy to akrasia just noted - why a more extended investigation into the nature of emotions than the one found in 5.4.5 would further illuminate the theory of phenomenal reasons and reason-governed action developed in this work. First, emotions are presumably the mental phenomena most closely related to, and most liable to get mixed up with, attitudes of wanting. The differences need therefore be carefully pin-pointed. In particular, I

would like to develop my suggestion that, whereas the objects of wanting are states of affairs the realization of which is thought possibly to be within one's power, the objects of emotion are states of affairs of which the negation of this is true. However, the validity of this thesis must somehow, presumably, be restricted, for some emotions appear to be essentially directed not towards states of affairs but towards concrete objects (e.g. love of a person). A similarity of particular interest between wants and emotions is that it is possible to give reasons for both. A comparison between these types of reason would be enlightening. Secondly, a firmer grasp on emotion would be conducive to a firmer grasp on the non-intentional character of the behaviour resulting from emotion, which in its turn would deepen the understanding of intentional action by placing it against a contrasting background.

(5) On my reductive account, action-theoretical concepts like wanting, having a reason, acting on a reason, acting at will and trying are analyzed in terms of propositional thinking, logical and evidential relations between the contents of such thoughts and the causal power of such thinking. I have tried to say something informative about these matters, about propositional thinking, conditionals and causation, but it is apparent that most of what ideally should be said has been left unsaid. What I should like to claim is to have brought action theory to a point where further progress in understanding is the responsibility of other disciplines, viz. the theory of theoretical thinking (epistemology) and the theory of causation.

NOTES

Introduction

1. Cf. [15], p. 182. It is another question whether Austin in practice lived up to this programmatic remark: a reader of e.g. [17] may be tempted to answer "no".
2. [129], p. 117. Manser also sees a discrepancy between Austin's programme and practice, but regards the latter as the more insightful.

Chapter 1

1. [24], chap. 6.2. Recently, in [25], Baier has modified his position in my direction: here he recognizes only two types of reason, explanatory and directive ones.
2. Another difference Baier claims to find is that only explanatory reasons have motivational power, but from this I dissent, as will become apparent in 5.1. The resemblance noted in the text can only be rather superficial, since Baier's conception of reasons differs radically from mine (see 5.1).
3. This distinction also goes back to Baier [24], e.g. p. 157.
4. I will in chap. 2 defend the thesis that acting can be analyzed as causing something to become a fact.
5. James Rachels has suggested [167], roughly, that it is a proposition backing the hypothetical, rather than the hypothetical itself, that is a reason, e.g. that someone's reason for rushing out of a theatre is not that (only) if he does so will he avoid being burned, but that the theatre is on fire. But surely if one knows merely that the conditional holds - without knowing why this is so - then this is sufficient to provide one with a reason for rushing out. (A further inadequacy in Rachels' treatment is that he demands that the reason-constituting beliefs be true, that they amount to knowledge. His proposal (p. 180) that somebody whose motivating beliefs are false merely thinks he has got a reason is, however, unsatisfactory, since it is by no means necessary that an agent be aware of the beliefs motivating him as beliefs motivating him, i.e. as reasons.)

Chapter 2

1. For instance, both Davidson's ([66], pp. 6-8) and Goldman's ([88], pp. 17-8, 44-8, 63-72) conceptions of action are subject to this stricture, since they require that something must be intentionally done in order for an action to have occurred (and the intentionality of an action derives, according to them, from its being caused by reasons). Don Locke's ([122], pp. 30-1) view of action - which lays down that a bodily movement is an action just in case it takes place under control, i.e. just in case the agent's decision or intention to stop or modify the movement would make the movement conform - is likewise vulnerable to this objection.
2. This appears to be Davidson's standpoint, because of (a) his afore-mentioned conception of action, and (b) his view that e.g. beliefs cannot be ascribed to non-language users [69].
3. Amongst others, Goldman, since though he agrees with (a), he dissents from (b) in note 2.
4. [40], chaps. 2 & 4; [103]; [112], p. 46; [202], pp. 252-4 and [208], esp. p. 406.
5. [216], pp. 39-41. For the addition of "intrinsic", see [195].
6. Cf. Kenny's distinction between "performances" and "activities" in [111], chap. 8.
7. Though I suppress the distinction in this work, just as I neglect the

- nuances that keep "act" and "action" apart.
8. J.S. Mill was a champion of this sort of position:
 "/An action is/ (n)ot one thing, but a series of two things: the state of mind called volition, followed by an effect ... the two together constitute the action" ([139], l, iii. 5);
 though his action-concept is obviously narrower than mine. More recently, Thomson ([202], chap. XVIII) has argued that an action is a causing or a fusion of causings.
 9. [27], p. 163. Perhaps the identification of a bodily event and a bodily action is even more clearly visible in this passage:
 "/Brown's/ hand's moving is the same event as his hand's being moved by Brown" ([26], p. 266).
 10. [66], p. 23. Perhaps Davidson is also inclined to take Beardsley's line of identifying somebody's moving his body with his body moving (see e.g. [61], p. 116).
 11. Chisholm appears to be most uncompromising in [49], pp. 17-22. Later, in [53], he seems to take a less extreme stand. [199], chap. 8-9.
 12. [217], p. 72; cp. [218], esp. pp. 48-54. Von Wright's concept of action is more restricted than the one defined in this chapter since it includes a reference to intentionality.
 13. The argument is found in [217], pp. 60-82 and [218], Pt. II.
 14. So far as I can find, the two quotations in the preceding paragraph do not make quite the same point, for one might know in general that there is a cause without having knowledge of what the cause is. Von Wright's claim is, I think, tenable only if it excludes no more than the second form of knowledge.
 15. I think Mackie is prone to handle over-determination in the same way that I do (cp. [128], p. 47).
 16. This stipulation brings my definition of inanimate action close to Brown's ([40], p. 139), since he also requires the cause to be a "moving and changing". Brown, however, does not accept a causal analysis of human action, apparently to a large extent because he rejects out of hand the idea that mental episodes can cause bodily movements (e.g. pp. 45-8, 120).
 17. Cf. "I cut my hair" which can be used even when I let a barber do it ([208], p. 406). Thomson dismisses the use of "killed" in the text as "special" without any argument ([202], p. 128). It is significant, I think, that those who deny that causing to die is sufficient for killing have "no even remotely plausible answer" ([202], p. 130) to what is further required.
 18. For an analysis of prevention, see [35].
 19. Much of the impetus of the recent discussion of this topic derives from Danto's [56] and [57]; see also his [59].
 20. It is difficult to find a passage that states this expressly, but in [59], p. 63 a basic action is identified with a physiological process, and such a process occurs, of course, within the confines of the body. Cf. also [59], p. 141.
 21. Cf. [26], p. 270 where this difficulty is dealt with in a way which seems similar to mine.
 22. See [79]; actually Fodor says that the latter sentence is ungrammatical, but this seems false, for it may be true when e.g. the stab puts a in a state of shock and confusion which leads him to kill b on Sunday.
 23. Complex actions have been discussed e.g. in [88], p. 22 and [59], p. 29.

Chapter 3

1. Sometimes (e.g. in [206] and [84], pp. 231-44) it is argued that sentences like "I think (believe etc.) p", in at least some of their uses, are not employed to make psychological statements, but to put forward p in a

cautious fashion. I think this view arises from the false assumption that these two functions exclude each other. In fact, given the reporting use, the other can be construed as an additional implication in certain contexts (cf. [87], chap. III).

Hunter makes a similar mistake about "I intend to do ..." when he states that the function it is endowed with is not that of reporting a mental state, but of "authorizing an expectation as to our doing something" ([106], p. 43).

2. "An idea assented to feels different from a fictitious idea, that fancy alone presents to us: And this different feeling I endeavour to explain by calling it a superior force, or vivacity ..." ([105], p. 146).
3. Vendler is aware of this limitation on substitutivity, but thinks - for reasons that escape me - that only concessions, not total surrender, on his part are necessitated by it ([207], pp. 69-82).
4. For further amplification of this point, see [80], chap. 1.
5. Cf. e.g. [180], p. 152, and also [187], where an analogy similar to that to follow in the text is found.
6. For a similar idea, though physiologically slanted, vide [77], pp. 16-7. The notion of explicit entailment will be further discussed in 3.2.2.
7. Cf. Vendler's distinction between "saying" in a strong and weak sense, [207], pp. 25-6, 53-4, 61-3.
8. To my knowledge, the terms "belief de re" and "belief de dicto" were introduced by Sosa in [191]. It is not clear whether Sosa views his distinction as a distinction between senses of "believe", but Chisholm appears to do so in [54], pp. 159-60.
9. Cf. this account with Beebe's distinction between referential and attributive propositional attitudes in [28].
10. Cf. Burge's criticism of substitutivity as a criterion of belief de re, [41], pp. 341-2.
11. From this it follows that I am unmoved by Kripke's attempt to extend his argument to terms for natural kinds, e.g. "rabbit" and "lapin", pp. 264-5.
12. This example figures in a similar environment in [75], sec. 3.17.
13. Edgley's concept of a normative principle is weaker than mine in that he does not require that it must be possible consciously not to conform to it, only that non-conformity simpliciter must be possible ([75], p. 74). Thus, he can consistently argue (secs. 3.14-3.18) that rules like "It is wrong to believe what is contradictory" are normative (since people often believe what, unbeknownst to them, are contradictions), though, on the whole, he agrees with me in regarding it as impossible to believe what is an explicit contradiction. A disadvantage with Edgley's terminology is that paradigmatic normative rules - like moral or legal principles - are normative in my stronger sense. The difference between them and the law of contradiction is consequently left unmarked by Edgley and thus runs the risk of being overlooked.
14. If a distinction between the causal theory and the functionalism of e.g. Putnam ([163], chaps. 14, 16, 18, 20, 21) and Fodor ([78], chaps. 3 & 4) is desired, it appears reasonable to draw it in the way suggested by Levin ([119], pp. 105, 129): functionalism does not specify the role as a causal one, but leaves it uninterpreted - although such a demarcation does not square with everything e.g. Fodor says ([78], pp. 107-9). But I need not bother to stress the distinction in this connection.
15. The phrase "apt for bringing about" will not bear exactly the same sense in the analysis of every mental concept - a case in point is belief and will, [5], pp. 82-3, 249.
16. By "behaviour" Armstrong must mean bodily movements - otherwise a charge of circularity might arise (cf. [5], p. 84).
17. See e.g. [120]; [121]; [100], chaps. 3 & 4; [78], p. 143; [190].
18. Armstrong criticizes Braithwaite but his criticism - that Braithwaite

- overlooks that "belief" designates a causally active factor [6], pp. 71-2) - seems unfair.
19. Ryle seems half-aware of this: at one point, he adds "and also cares whether it is thin or thick" ([172], p. 135, my emphasis).
 20. The idea that mental concepts are comparable to theoretical concepts was (to my knowledge) first advanced by Wilfrid Sellars in [179]. But for Sellars it is only "thought" as applied to inner episodes that is a theoretical construct; he also recognizes a behavioural construing of it, which serves as a model for the theoretical employment.
 21. Sometimes, however, a topic is felt to be so charged that a person is unable to bring himself to consider it in this reflective fashion; whenever the question comes up, he blocks consideration of reasons by stubbornly repeating to himself that he has a certain belief or that something is true - thus becoming perhaps a victim of self-deception.
 22. The Armstrong-Braithwaite view is also vulnerable to this objection, as pointed out in [157], pp. 128-9; cf. also [161], pp. 248-50, 265.
 23. Dennett's idea of an "intentional system", of replacing the question whether an organism really has beliefs and desires with the question whether the ascription of beliefs and desires to the organism makes possible effective explanation and prediction of the organism's behaviour, of replacing a question concerning what is true or false with one concerning what is pragmatically justifiable (cf. [73], e.g. p. 7), is infected with a similar circularity: the notions of explanation and prediction make sense only among creatures that have beliefs, and their having beliefs cannot without vicious circularity be elucidated in Dennett's fashion.
 24. Relevant works of Sellars' include: [178]; [179]; [182], chaps. 3 & 4, [183]; [187] and [47]. Aune: [11], chap. 8; cf. also Tuomela [205], chap. 3. A similar, but less developed theory is found in [86].
 25. This has profited from Marras' criticism in [130], [131], [132] and [133], even though it may not be identical in every detail to anything found in these papers.
 26. These ways out are suggested by Tuomela ([205], pp. 55, 61-2), although he does not seem altogether pleased with them, but is rather inclined to give up the project of analyzing conceptuality strictly in terms of rules of language (pp. 59-62). It is not clear to me what Tuomela wants to substitute for this appeal to linguistic rules or what he proposes to complement it with. His "'public' and 'social' view of inner mental episodes" (p. 63) is hardly intended to be entirely behaviouristic.
 27. Cf. the thought/language view endorsed by Chisholm in [47], pp. 524, 533. Cf. also H.P. Grice's view of meaning in [92] which has been developed e.g. in [29].
 28. The introduction of isomorphic symbols necessitates a modification of (T6) in 3.2.1: "translation" must be taken to include the replacement of images by linguistic units corresponding to them.
 29. Fodor seems to lean towards such a view ([80], pp. 184-7). See also his defence of images against some philosophical arguments (pp. 187-94). That apart, he is not very explicit about the nature of the language of thought. Sometimes he seems to imply that the language is not psychological at all (e.g. the "language ... /the/ ... nervous system uses for computing", p. 68).

Chapter 4

1. I do not intend (C1) to cover cases where "if" is merely shorthand for "even if".
2. The term, but not the concept, derives from [218], p. 21; the concept has a considerably longer history.

3. John Young, another foe to the reduction of conditionals to material implication, is also willing to concede the consistency of "If p , q " and "If p , $\neg q$ " in reductio-contexts ([220], p. 60). But he also points to (p. 59) what is perhaps the gravest difficulty for the material implication approach: that it is extremely hard to explain how " $\neg$ (If p , q)" and " p & $\neg q$ " could be equivalent.
4. This sentence should be read as saying that a , at t , has the ability and opportunity to lift the stone at t . Contrary to what some philosophers ([88], p. 204-7; [118], e.g. p. 243) seem to have supposed, one cannot prior to t have the ability-cum-opportunity to lift the stone at t , though one can then have the ability-cum-opportunity to cause that one at t lifts - or has the ability-cum-opportunity to lift - the stone.
5. I take it for granted that the want mentioned in the antecedent is stronger than every other incompatible want the agent has at the relevant time.
6. Ayers considers (pp. 134-5) a countermove of this sort - which would make the analysis of one concept of "personal power" parallel to Ayers' of "natural power" (for the latter, see [18], e.g. p. 69) - but rejects it mainly, it seems, because he is unaware of the distinction between (a) and (b)-types of abilities.
7. We must not confuse knowledge how to do something with knowledge how something is done, in the sense of knowledge how an action is constituted in finer detail. Of course, bodily actions are - as indicated in the text - the proper object of the latter form of knowledge-how. My argument is that the possession of the latter form of knowledge-how is not sufficient for possession of the former; a further precondition is that the latter knowledge informs one of something that can be done as a means.
8. Cf. the latter reading of "can ... at will" with Goldman's epistemic sense of "ability" ([88], p. 203). The domain of applicability of the two senses of "can ... at will" will be further discussed in 6.4.

Chapter 5

1. What are real reasons for a person equally depends on his wants, but since phenomenal reasons are of special interest here, I formulate the conditions in terms of such reasons.
2. Baier himself seems committed to a similar interpretation, namely in terms of "best reasons" ([24], p. 86-88). The details of the interpretation do not matter so long as it is agreed that an allusion to reasons figures in it.
3. Baier's animosity towards the idea that reasons are dependent on wants or desires may to a large extent stem from his crude conception of a desire or a want (in one sense of the word "want", whilst in another he thinks it signifies a disposition to make a request ([24], pp. 114-5)): he seems to think that a desire involves "a feeling of restlessness" (p. 110) or "a felt impulse toward doing something" (p. 111). Of course, it is grotesque to tie reasons to desires thus conceived.
4. [159], pp. 190-3. In further conformity with my view, Phillips Griffiths regards ([159], pp. 189-90) the contents of beliefs as "assessable reasons". Cf. also Locke's view on these two matters in [123]. A difference between these authors and myself is that we approach the relation between wanting and having a reason from opposite directions: they explicate wanting to do F in terms of having a reason - in Locke's case a reason which is effective as a reason ([123], p. 174) - to do F , while my analysis goes in the reverse direction. Phillips Griffiths^T and Locke's view has, in my eyes, the strained consequence that they cannot contrast doing something (because one wants it) for its own sake with doing it for a reason.

5. Vide esp [60], pp. 79-84, but also e.g. [68], pp. 147-8, and [70], p. 241. Davidson sometimes uses "pro-attitude" to designate wanting in the broad sense.
6. Cf. e.g. [85], p. 29 and [136]:
 "what I say when I give you my reason must make clear both my desire and belief ... That is a condition for the success of an explanation of some-one's action in terms of his reason for doing it" (p. 53, my emphasis).
7. Hence, contrary to what is said e.g. in [60], p. 82, [169], p. 181 and [95], p. 239, wanting to F can be no reason for doing F, since it is obviously no reason for wanting to F.
8. The wanting here is wanting_d in the terminology introduced in 5.4.1.
9. Gilbert Harman asserts that intentions are self-referential, that "the intention to do A contains the intention of intending to do A" ([101], p. 441). Harman denies (p. 440) that an infinite regress lurks here, claiming that the intentions generated will not be "distinct". However, his doctrine (pp. 438-41) that forming an intention to do A is a means of doing A seems at odds with this contention, for an intention to do A must be distinct from (less specified than) an intention to do A by means of doing B, even if doing B is a logically necessary means, since the context here is intensional. Therefore, on Harman's account, having an intention is subject to the absurdity of involving the having an infinite number of intentions. I will, in "Conclusions and Limitations" under (1), attempt to demolish Harman's main argument for the view that intentions are self-referential.
10. A denial of (1) has recently come to my attention. William Alston postulates - "Self-intervention and the Structure of Motivation" in [141] - two separate, hierarchically ordered motivational systems, each of which is influenced by its own characteristic considerations: a higher-level system influenced by considerations of principle and a lower-level one influenced by hedonic considerations (see pp. 86ff). I think there are decisive objections to this idea. First, with the specification of the two types of motivating considerations, for (a) the independency of Alston's two types seems questionable, and (b) it is even more questionable whether considerations of principle are primarily only second-order incentives to rearrange wants, not also first-order ones to act in conformity with the principles. Secondly, if two motivational systems are postulated, why not three, four ...? As Alston admits (pp. 83-4), you can want to reform your second-order wants as well. This magnifies the first problem.
11. [162], pp. 62-4; though Prichard here speaks of "willings", i.e. acts of will. Cf. also Meiland on the object of intention in [137], pp. 39-43.
12. [20], p. 658. (But in [23] she rejects a corresponding thesis for wanting.) Baier criticizes ([20], pp. 652ff) the view (which she attributes, perhaps incorrectly, to Chisholm [51]) that intending to do is intending to bring about that one does something. But her discussion of the relationship between "intending to do" and "intending to bring about" is flawed by: (i) her negligence of the possibility that intending to do is intending to bring about, not an action, but its result, (ii) her assumption (pp. 653-5) that "bring about" marks the applicability of the "how" of method, i.e. that bringing about is achieving something indirectly. But bringing about is (at least for me) a purely causal notion, and there is direct causation as well as indirect.
13. Hume classifies desire and aversion as (direct) passions ([105], Bk. II Pt. III, sec. I). He appears to conceive a passion on the model of a sensation, for he asserts (*ibid.*, sec. III) that it "contains not any representative quality". Desire and aversion seem to be seen as occurring prior to the action and as mediating between it and "the prospect of pain or pleasure":

- "when we have the prospect of pain or pleasure from any object, we feel a consequent emotion of aversion or propensity, and are carry'd to avoid or embrace what will give us this uneasiness or satisfaction" (*ibid.*, sec. III, p. 461).
14. [16], pp. 226-7; I have adjusted and simplified his exposition.
 15. A distinction closely resembling mine is found in e.g. [34], p. 945, [142], secs. 2 & 3 and [189], p. 177.
 16. I take an instant to be an interval of so short a duration that a human being cannot perceive any change within it. If this is so, one cannot have a want that lasts only an instant (given that a want is a disposition). For this reason, the phrase "begins to want at t" is preferable to "wants at t", but for the sake of simplicity I shall employ the latter.
 17. These two objections are suggested by remarks of Radford in [168], pp. 53-4.
 18. Someone agreeing with Hornsby that "(s)ince wanting moves nothing, something must intervene between your wanting and your finger's moving" ([104], p. 56), will perhaps dispute that a propositional thought can elicit the bodily movement in an intentional action - perhaps something in the "executive mode" is felt to be missing. But we have seen here that propositional thoughts are capable of causing reflex-behaviour. The propositional content of these causally potent thoughts may well be the same as the propositions that form the core of the content of thoughts causally active in intentional action. The feature of intentionality only originates from the embedding of these propositions in the mode of the attainable.
 19. This should suffice to refute Nowell-Smith's suggestion ([148], pp. 113-4), taken up by Abelson ([1], pp. 48-53), that "because I want to" is a "reason-terminating" locution, i.e. a locution that indicates that the chain of reasons has reached its end.
 20. For a classic defence of this position, see Davidson's [60].
 21. This form of compromise between causalism and non-causalism is also found in [167], pp. 183-4 and in [123], p. 177.
 22. One pertinent formulation of Kant's is:
 "Wer den Zweck will, will (sofern die Vernunft auf seine Handlungen entscheidenden Einfluss hat) auch das dazu unentbehrlich notwendige Mittel, das in seiner Gewalt ist. Dieser Satz ist, was das Wollen betrifft, analytisch ..." ([109], p. 417, Acad. Ed.)
 It is debatable whether Kant's view is that wanting the necessary means is something logically necessary given the end and the means-thought, or that it is only logically necessary given this and conformity with the norm of rationality. For a discussion that favours the latter alternative, see [45], pp. 300-3. In 6.1.1 I will argue that, whatever Kant's view is, the former is correct.
 23. The inspiration for treating "cancellability" as a sign of an implication being "pragmatic" or "contextual" comes from [93], pp. 90-5.
 24. For other champions of the same view, see the discussion of the theoretical construct approach in 5.4.5.

Goldman argues ([88], pp. 112-4) correctly that there is no incompatibility between wants causing behaviour and their being logically tied to it through being defined as potential causes of behaviour. But he mixes up this thesis with the falsehood that descriptions like "want" are both eligible for inclusion in causal laws of behaviour and logically connected to behaviour. The latter thesis can consistently be held only if one loads "want" with more than one sense. This is precisely what Goldman does: in his exposition the term vacillates between "thought of an optative content", (see esp. pp. 94-105), which is a purely internal characterization of a thought, and "mental event/presumably a thought of the afore-mentioned kind/ that tends to result in behavior" (p. 112),

which is a partly functional or relational description.

25. Perhaps Lennart Nordenfelt's position ([145], chap. 3) is the one most closely resembling mine, though he uses "intention" roughly where I would like to see "wanting" and makes the latter dependent on the former rather than the other way around. Some philosophers, who hold that wants, though causes of behaviour, do not, thus described, enter into the statement of strict laws, take the same view of other propositional attitudes, e.g. Davidson in [67].

To some extent, my theory is foreshadowed in Schopenhauer's writings on the topic (e.g. [174], §§ 17-8 & [175], pp. 35ff). Schopenhauer clearly saw that an appeal to the representations of our faculty of knowledge was not enough to explain human actions. Therefore, he postulated an individual, objectless will (or character) which together with representations causes specific actions or acts of will. In this, Schopenhauer thought, the explanation of action is just like explanations of natural phenomena: the latter all presuppose some natural power, like gravitation, impenetrability, or resilience, which is triggered off by a stimulus.

26. Mackie raised, if I remember correctly, the following objection in a conversation: how could a statement, which is conclusively verified by a perceptual experience, say something about a counterfactual situation? It would seem that such a statement must restrict itself to the situation the experience is an experience of. Powerful as this objection may seem, I still find it untenable. Note that there are further counter-examples to it. An identity-statement (e.g. that the man dressed in black is identical with the man smoking a pipe) can be verified by observation, although it carries counterfactual implications (e.g. that if the man dressed black had been somewhere else, so would the man smoking a pipe). It is tempting, but wrong, to assume that if p entails q , q cannot contain any ideas over and above what p contains; for p entails p or r , even though p does not contain the ideas turning up in r . An attempt might be made to rescue the principle by restricting it to entailments constituting the analysis of a statement - p or r cannot figure in the analysis of p , since that would be blatantly circular. But I have not committed myself to the view that causal assertions can be analysed in counterfactual terms. I am, on the contrary, inclined to believe that such an analysis would be circular - that the notion of causation is needed when the precise force of the relevant counterfactuals is spelt out. The reflections in the text suggest that "causes" might well designate an unanalyzable relation we are experientially acquainted with (just as "red" designate such a property).

27. What has been asserted in this and the preceding paragraph should be compared to Armstrong's view in [5], pp. 97-8. (Cf. also [213], pp. 170-4, 180-2.) But Armstrong is prepared to go a step further than me, to claim that there can be perception and introspection of something having only relational properties, of a causal relation between phenomena which are not in the perceptual experience given any internal properties. This I am unable to accept; in his example of the pressure in the back (p. 97) I would say that we must feel at least the rough shape and size of that which exerts the pressure.

28. Already Anscombe has observed ([4], p. 15) that we have knowledge "without observation" (as she calls it) of causes in reflex-action. But she seems to take this as evidence that there is a special type of "mental" causation ([4], p. 16) at work here. My earlier examples should knock away the support for her contention that the causation immediately known is peculiar to the mental sphere.

Chapter 6

1. A proviso is added to this in 6.3.2.
2. [75], p. 121; the validity of the quoted principle is, however, restricted to "central cases of belief".
3. Edgley's point is not that one may fail to do y through inability: the table on p. 145 makes it plain that "may not will y" can be substituted for "may not do y". - The other way alluded to in which reason may fail to decide one's conduct is the trivial one - when one does not realize that y is necessary for x ([75], pp. 143-4).
4. Here the stricture on second-order desires (see 5.2.2) is relevant again, since a desire to be rational (in the practical dimension) is a desire to desire certain things in certain circumstances.
5. In this definition, "knew" and "wanted" are supposed to be used in an occurrent sense, not as designating dispositions, cf. [27], p. 167. For a similar definition, see [95], pp. 246-8.
6. This does not prevent (a) in Beardsley's definition from being true, that is, for it to be true that the arm's going up is describable as a (bodily) doing of a's, in the sense fixed by Beardsley ([27], pp. 163-4), since the proximate cause of the movement is muscular changes in a's body (cp. p. 16).
7. Similar counter-instances have been presented earlier by other philosophers - e.g. Taylor ([199], pp. 248ff) - but less compellingly.
8. Pears' proposal - that the agent must intend the causal connection between an act of will and a fulfilling result to be of the "reliable" and "normal" type ([156], p. 68) - is open to a similar criticism: at least in the case of some acts, a connection involving nervous excitement might be both reliable and normal. For further criticism of Pears, see [151], pp. 89-95.
9. Goldman commits himself to such mental states in [89], pp. 70-1; but here he also tries to be a bit more specific about the feeling (pp. 76-82).
10. [83], p. 158. In Frankfurt's terminology, "action" includes reference to intentionality; it is intentional movement ([80], p. 159).
11. Cf. R. Martin's criticism of Stoutland in [135], pp. 341-6.
12. E.g. O'Shaughnessy in [150], pp. 365ff, McCann in [124], pp. 424-7, Hornsby [104], pp. 34-6, and Davis [72], pp. 16-7.
13. McCullagh [126] asserts that there must not be a sufficient psychological or physiological condition at the time of acting apart from the want and the physiological state correlated with it. If by "at the time of doing it" ([126], p. 202) he means throughout the time doing it, he makes a demand akin to that in the text. Let me remind the reader that as I use "cause" it entails "non-redundant part of every sufficient condition that is present" (see pp. 23-4); so causal terminology in my approaching definition of acting intentionally ensures non-redundancy.
14. Inspiration for a distinction between ontological and teleological basicity originally comes from [21], p. 168.
15. The conclusion Danto draws from this - that "with regard to basic actions, trying them consists in doing them, albeit strenuously" ([58], p. 59), that an attempt to do a basic action cannot be unsuccessful - does not follow, for although no means are applied in the execution of a teleologically basic action, it is still not ontologically simple and basic (in an absolute sense). Consequently, an attempt at a basic action might be unsuccessful in that it consists in an action ontologically simpler or more basic than the one attempted (e.g. an attempt to raise an arm results only in slight muscular contractions).
16. I use "non-intentionally" rather than "unintentionally", because I regard it as natural to reserve the latter for situations in which the agent fails to realize what he wants and instead realizes something incompatible with this.

17. [4], § 48. Cf. also J.R.S. Wilson [214], pp. 115-6, where a similar thesis is put forward for basic action. Wilson seems to have a sounder conception of non-observational knowledge than Anscombe.
18. Cf. [107], pp. 487-8. Note that acting at will is not logically posterior to acting on reflex, since, conceivably, the images could be innate. - James' "ideo-motor" theory of voluntary action (propounded in [107], chap. 26) is an ancestor of Bach's theory. James' theory has recently been revived also in psychological circles; vide e.g. [91] and [113].
19. Nor does Bach restrict his theory to basic action.
20. Cf. how Castañeda couches his "practitions" ([45], e.g. pp. 169-72).
21. Such a view could be read into [89], pp. 79-80, especially if the passage is read in conjunction with the claim that "to have a want is to 'assent' to an optative proposition" ([88], p. 102). James accords ([107], pp. 568-9) a considerably more marginal role to such a practical modality.
22. Consequently, volitions postulated ([72], pp. 16, 20-1) as the causes of the correct views are superfluous.
23. Conditions (6), (7) and (9) are not entirely accurate when *r* is the result of a consequential action, i.e. a state of affairs whose existence is so indirectly caused by facts of change pertaining to the agent that nothing causally relevant need be true of the agent at the time *r* takes place. Suppose I set a boulder in motion for the purpose of destroying a cabin and its inhabitants in the valley below, but that I am seized by compassion and change my wants after having set it in motion but before it reaches the hut; then this does not prevent it from being true that I intentionally destroyed the hut (if that is the result of the impact of the boulder). For a consequential action to be intentional, it is sufficient that (6), (7) and (9) hold for the duration of the essential act that generates it. In opposition to this, Beardsley maintains ([27], pp. 168-9) that an action is intentional only as long as the appropriate beliefs and wants obtain. This has counter-intuitive consequences, e.g. that it becomes logically impossible intentionally to cause grief to people who love one by committing suicide.
24. This is a much-debated question; for further references and a stand opposite to mine, see [140].
25. Examples of this kind were perhaps first discussed by Taylor ([199], pp. 194-5) and Chisholm ([49], p. 44).
26. This is a further difference - and the reverse of the one noted on pp. 92-3 - between theoretical and practical reasoning, since if *a* believes that if *p* then *q* and that *p*, he must believe that *q*.
27. I think a corresponding thesis holds for theoretical reasons: *p* is a reason for believing *q* only if the antecedent of the conditional "If *p*, *q*" is verifiable independently of the consequent. "Reason for" is an epistemic relation.
28. This distinction is borrowed from [134], p. 60.
29. Cf. [113], but contrast [89], pp. 72-5, where it is asserted that these component-actions are voluntary (though not intentional). But the rigidity and inflexibility of the ordering of these components, the difficulty of changing the order at will (e.g. changing the word-order of a poem learnt by heart) certainly suggest independence of the will.
30. Intuitions seem to diverge here: e.g. Bach demands knowledge ([19], pp. 375-6), while Pears seems to be on my side ([156], p. 67).

Conclusions and Limitations

1. For further criticism of Castañeda on this point, see [96], pp. 266-8.
2. It might be retorted that though this argument is effective against analogy theorists like Sellars and Aune, it carries no force against someone who, like myself, is prepared to countenance a natural medium of

thinking-that. However, nothing adduced in my vindication of such a medium requires it to be rich enough to express non-propositional that-clauses.

3. Cf. [97], p. 97:
"the ... word 'intention', since it is associated with the possibility of declaration of intention, is out of place in the context of animal behaviour".
But should anyone stubbornly insist on the superfluity of reference to shall-thoughts, I would recant: the essential point is that intentions are not sources of motivation independent of wanting.
4. Wishing is analogous to intending and deciding: a wish is the expression (in e.g. "If only p" or "Would that p") of an emotion (of hoping that p, being sad that -p etc.).
5. I have throughout employed "intentionally" and "voluntarily" (or "at will") as synonyms, because I have assumed that an action can be intentional without an intention being present among its antecedents. (Cf. [9], p. 401; [71], p. 46; [72], pp. 59-60; [101], p. 434.) My reason for this is not that I am convinced that this accords with non-technical usage, but that I do not take intentions to be motivationally potent.
6. This view appears to have originated from [98], but it is also found e.g. in [9]; [27]; [72], pp. 72-82; [94]; [101]. Davidson argues against it (in [71], pp. 50-2), but his arguments do not sway me.
7. This appears to be Harman's major argument ([101], p. 442) for the view - discussed in note 9 to chap. 5 - that intentions are self-referential.
8. In two recent papers [10] and [209] a loosening of the tie between evaluations or reasons and (some) wants is advocated. In another two recent papers, [39] and [65], the solution espoused is, roughly, that of opening up a gap between wants based on reasons and an unconditional want or intention.

BIBLIOGRAPHY

- APQ = American Philosophical Quarterly
 AJP = Australasian Journal of Philosophy
 CJP = Canadian Journal of Philosophy
 JP = Journal of Philosophy
 PR = Philosophical Review
 PS = Philosophical Studies
 PPR = Philosophy and Phenomenological Research
 PAS = Proceedings of Aristotelian Society
 RM = Review of Metaphysics
- [1] R. Abelson: Persons, Macmillan, London & Basingstoke, 1977.
 - [2] W.P. Alston: "Wants, Actions and Causal Explanation" in [43].
 - [3] W.P. Alston: "Motives and Motivation" in P. Edwards (ed): Encyclopaedia of Philosophy, Macmillan, London & Basingstoke, 1967.
 - [4] G.E.M. Anscombe: Intention, Blackwell, Oxford, 1957.
 - [5] D.M. Armstrong: A Materialist Theory of the Mind, Routledge & Kegan Paul, London, 1968.
 - [6] D.M. Armstrong: Belief, Truth and Knowledge, Cambridge U.P., 1973.
 - [7] R. Audi: "The Concept of Believing", The Personalist, 53(1972), 43-62
 - [8] R. Audi: "The Concept of Wanting", PS, 24 (1973) 1-21.
 - [9] R. Audi: "Intending", JP, 70 (1973), 387-403.
 - [10] R. Audi: "Weakness of Will and Practical Judgement", Nous, 13(1979), 173-96.
 - [11] B. Aune: Knowledge, Mind and Nature, Random House, New York, 1967.
 - [12] B. Aune: "Hypotheticals and 'Can': Another look", Analysis, 27(1966-7), 191-5.
 - [13] B. Aune: Reason and Action, Reidel, Dordrecht, 1977.
 - [14] J.L. Austin: Philosophical Papers, Oxford U.P., 2ed., 1970.
 - [15] J.L. Austin: "A Plea for Excuses" in [14].
 - [16] J.L. Austin: "Ifs and Cans" in [14].
 - [17] J.L. Austin: Sense and Sensibilia, Clarendon Press, Oxford, 1961.
 - [18] M.R. Ayers: The Refutation of Determinism, Methuen, London, 1968.
 - [19] K. Bach: "A Representational Theory of Action", PS, 34(1978), 361-79.
 - [20] A. Baier: "Act and Intent", JP, 67(1970), 648-58.
 - [21] A. Baier: "The Search for Basic Actions", APQ, 8(1971), 161-70.
 - [22] A. Baier: "Intention, Practical Knowledge and Representation" in (36).
 - [23] A. Baier: "Mixing Memory and Desire", APQ, 13(1976), 213-20.
 - [24] K. Baier: Moral Point of View, Cornell, U.P., Ithaca & London, 1958.
 - [25] K. Baier: "Moral Reasons", Midwest Studies in Philosophy, 3(1978), 62-74.
 - [26] M.C. Beardsley: "Actions and Events: The Problem of Individuation", APQ, 12(1975), 263-76.
 - [27] M.C. Beardsley: "Intending" in A.I. Goldman & J. Kim: Values and Morals, Reidel, Dordrecht, 1978.
 - [28] M. Beebe: "How Beliefs Find their Objects", CJP, 9(1979), 595-608.
 - [29] J. Bennett: Linguistic Behaviour, Cambridge U.P., 1976.
 - [30] R. Binkley et al (eds): Agent, Action, and Reason, Blackwell, Oxford, 1971.
 - [31] R.B. Braithwaite: "The Nature of Believing" repr. in [158].
 - [32] M. Brand (ed): The Nature of Human Action, Scott, Foresman & Co., Glenview, Illinois, 1970.
 - [33] M. Brand: "Introduction" to Sec. II of [32].
 - [34] M. Brand: "Causes of Action", JP, 67(1970), 932-47.
 - [35] M. Brand: "The Language of Not Doing", APQ, 8(1971), 45-53.
 - [36] M. Brand & D. Walton (eds): Action Theory, Reidel, Dordrecht, 1976.
 - [37] M. Brand: "The Fundamental Question in Action Theory", Nous, 13(1979), 131-51.
 - [38] R. Brandt & J. Kim: "Wants as Explanations of Actions" repr. in [42].

- [39] M. Bratman: "Practical Reasoning and Weakness of the Will", Nous, 13(1979), 153-71.
- [40] D.G. Brown: Action, University of Toronto Press, 1968.
- [41] T. Burge: Belief De Re, JP, 74(1977), 338-62.
- [42] N. Care & C. Landesman (eds): Readings in the Theory of Action, Indiana U.P., Bloomington, 1968.
- [43] H-N. Castañeda (ed): Intentionality, Minds and Perception, Wayne State U.P., Detroit, 1967.
- [44] H-N. Castañeda: "Intentions and Intending", APQ, 9(1972), 139-49.
- [45] H-N. Castañeda: Thinking and Doing, Reidel, Dordrecht, 1975.
- [46] R.M. Chisholm: "The Contrary-to-fact Conditional", Mind, 55(1946), 287-307.
- [47] The Chisholm-Sellars Correspondence on Intentionality, Minnesota Studies in The Philosophy of Science, Vol. II, University of Minnesota Press, 1958.
- [48] R.M. Chisholm: "J.L. Austin's Philosophical Papers" repr. in [76].
- [49] R.M. Chisholm: "Freedom and Action" in [117].
- [50] R.M. Chisholm: "He Could Have Done Otherwise", JP, 64(1967), 409-17.
- [51] R.M. Chisholm: "The Structure of Intention", JP, 67(1970), 633-48.
- [52] R.M. Chisholm: "The Logic of Intentional Action" in [30].
- [53] R.M. Chisholm: "The Agent as Cause" in [36].
- [54] R.M. Chisholm: Person and Object, Allen & Unwin, London, 1976.
- [55] P.M. Churchland: "The Logical Character of Action Explanations", PR 79(1970), 214-36.
- [56] A.C. Danto: "Basic Actions" repr. in [42].
- [57] A.C. Danto: "What We Can Do" repr. in [42].
- [58] A.C. Danto: "Freedom and Forbearance" in [117].
- [59] A.C. Danto: Analytical Philosophy of Action, Cambridge U.P., 1973
- [60] D. Davidson: "Actions, Reasons, and Causes" repr. in [210].
- [61] D. Davidson: "The Logical Form of Action Sentences" in N. Rescher (ed): The Logic of Decision and Action, Pittsburgh U.P., 1967.
- [62] D. Davidson: "Causal Relations" repr. in [192].
- [63] D. Davidson & J. Hintikka (eds): Words and Objections, Reidel, Dordrecht, 1969.
- [64] D. Davidson: "On Saying That" in [63].
- [65] D. Davidson: "How is Weakness of the Will Possible?" in J. Feinberg (ed): Moral Concepts, Oxford U.P., 1969.
- [66] D. Davidson: "Agency" in [30].
- [67] D. Davidson: "Mental Events" in L. Foster & J.W. Swanson (eds): Experience & Theory, Duckworth, London, 1971.
- [68] D. Davidson: "Freedom to Act" in T. Honderich (ed): Essays on Freedom of Action, Routledge & Kegan Paul, 1973.
- [69] D. Davidson: "Thought and Talk" in S. Guttenplan (ed): Mind and Language, Oxford U.P., 1975.
- [70] D. Davidson: "Hempel on Explaining Action", Erkenntnis, 10(1976), 239-53.
- [71] D. Davidson: "Intending" in Y. Yovel (ed): Philosophy of History & Action, Reidel, Dordrecht, 1978.
- [72] L.H. Davis: Theory of Action, Prentice-Hall, Englewood Cliffs, N.J., 1979.
- [73] D. Dennett: Brainstorms, Bradford, Vermont, 1978.
- [74] R. Descartes: Meditations.
- [75] R. Edgley: Reason in Theory and Practice, Hutchinson, London, 1969.
- [76] K.T. Fann: Symposium on J.L. Austin, Routledge & Kegan Paul, London, 1969.
- [77] H.H. Field: "Mental Representation", Erkenntnis, 13(1978), 9-61.
- [78] J.A. Fodor: Psychological Explanation, Random House, New York, 1968.
- [79] J.A. Fodor: "Three Reasons For Not Deriving 'Kill' From 'Cause to Die'", Linguistic Inquiry, 1(1970), 429-38.

- [80] J.A. Fodor: The Language of Thought, Harvester Press, Hassocks, 1976.
- [81] R. Foley: "Deliberate Action", PR, 86(1977), 58-69.
- [82] H.G. Frankfurt: "Freedom of the Will and the Concept of a Person", JP, 68(1971), 5-20.
- [83] H.G. Frankfurt: "The Problem of Action", APQ, 15(1978), 157-62.
- [84] M. Furberg: Saying and Meaning, Blackwell, Oxford, 1971.
- [85] D.P. Gauthier: Practical Reasoning, Clarendon Press, Oxford, 1963.
- [86] P. Geach: Mental Acts, Routledge & Kegan Paul, London, 1957.
- [87] M. Ginsberg: Mind & Belief, Allen & Unwin, London, 1972.
- [88] A.I. Goldman: A Theory of Human Action, Princeton U.P., 1976.
- [89] A.I. Goldman: "The Volitional Theory Revisited" in [36].
- [90] N. Goodman: Fact, Fiction and Forecast, Athlone Press, London, 1954.
- [91] A.G. Greenwald: "Sensory Feedback Mechanisms in Performance Control: With Special Reference to the Ideo-motor Mechanism", Psychological Review, 77(1970), 73-101.
- [92] H.P. Grice: "Meaning", PR, 66(1957), 377-88.
- [93] H.P. Grice: "The Causal Theory of Perception", repr. in G.J. Warnock (ed): The Philosophy of Perception, Oxford U.P., 1967.
- [94] H.P. Grice: "Intention and Uncertainty" in Proceedings of the British Academy, Vol. 57, Oxford U.P., 1971.
- [95] R. Grimm: "Purposive Actions", PS, 38(1980), 235-59.
- [96] D.F. Gustafson: "Castañeda's Intentions: A Critical Study of Castañeda's Thinking and Doing", Synthese 44(1980), 247-84.
- [97] S. Hampshire: Thought and Action, Chatto & Windus, London, 1959.
- [98] S. Hampshire & H.L.A. Hart: "Decision, Intention and Certainty" repr. in H. Morris (ed): Freedom and Responsibility, Stanford U.P., 1961.
- [99] S. Hampshire: Freedom of the Individual, Chatto & Windus, London, 2ed., 1975.
- [100] G. Harman: Thought, Princeton U.P., 1973.
- [101] G. Harman: "Practical Reasoning", RM, 29(1976), 431-63.
- [102] J. Hintikka (ed): Essays on Wittgenstein in Honour of von Wright, North-Holland, 1977.
- [103] D. Hirschmann: "Inanimate Agency", PAS, 72(1971/2).
- [104] J. Hornsby: Actions, Routledge & Kegan Paul, 1980.
- [105] D. Hume: A Treatise of Human Nature, Penguin ed.
- [106] J.F.M. Hunter: Intending, Dalhousie U.P., 1978.
- [107] W. James: Principles of Psychology, Vol. II, Henry Holt & Co., 1890.
- [108] N. Jareborg: Handling och uppsåt, Nordstedts & Söner, Stockholm, 1969.
- [109] I. Kant: Grundlegung zur Metaphysik der Sitten, Felix Meiner, Hamburg, 1965.
- [110] D. Kaplan: "Quantifying In" [63].
- [111] A. Kenny: Action, Emotion and Will, Routledge & Kegan Paul, London, 1963.
- [112] A. Kenny: Will, Freedom and Power, Blackwell, Oxford, 1976.
- [113] G.A. Kimble & L.C. Perlmuter: "The Problem of Volition", Psychological Review, 77(1970), 361-84.
- [114] S.A. Kripke: "Naming and Necessity" in D. Davidson & G. Harman (eds): Semantics of Natural Language, Reidel, Dordrecht, 1972.
- [115] S.A. Kripke: "A Puzzle about Belief" in A. Margalit (ed): Meaning and Use, Reidel, Dordrecht, 1979.
- [116] K. Lehrer: "Ifs and Cans" in [32].
- [117] K. Lehrer (ed): Freedom and Determinism, Random House, New York, 1966.
- [118] K. Lehrer: "'Can' in Theory and Practice: A Possible Worlds Analysis" in [36].
- [119] M.E. Levin: Metaphysics & the Mind-body Problem, Oxford U.P., 1979.
- [120] D.K. Lewis: "An Argument for the Identity Theory", JP, 63(1966), 17-25.
- [121] D.K. Lewis: "Psychophysical and Theoretical Identifications", AJP,

- 50(1972), 249-58.
- [122] D. Locke: "Action, Movement and Neurophysiology", Inquiry, 17(1974), 23-42.
- [123] D. Locke: "Reasons, Wants, and Causes", APQ, 11(1974), 169-79.
- [124] H. McCann: "Trying, Paralysis, and Volition", RM, 28(1974), 423-42.
- [125] H. McCann: "Volition and Basic Action", PR, 83(1974), 451-73.
- [126] B.C. McCullagh: "Causal Theories of Action", PS, 27(1975), 201-9.
- [127] J.L. Mackie: Truth, Probability and Paradox, Oxford U.P., 1973.
- [128] J.L. Mackie: The Cement of the Universe, Oxford U.P., 1974.
- [129] A. Manser: "Austin's 'Linguistic Phenomenology'" in E. Pivcevic (ed): Phenomenology and Philosophical Understanding, Cambridge U.P., 1975.
- [130] A. Marras: "Sellars on Thought and Language", Nous, 7(1973), 152-63.
- [131] A. Marras: "On Sellars' Linguistic Theory of Conceptual Activity", CJP, 2(1973), 471-83.
- [132] A. Marras: "Reply to Sellars", CJP, 2(1973), 495-501.
- [133] A. Marras: "The Behaviorist Foundation of Sellars' Semantics", Dialogue, 16(1977), 664-75.
- [134] J.B. Martin: "Basic Actions and Simple Actions", APQ, 9(1972), 59-68.
- [135] R. Martin: "The Problem of the 'Tie' in von Wright's Schema of Practical Inference: A Wittgensteinian Solution" in [102].
- [136] S. Meikle: "Reasons for Action", The Philosophical Quarterly, 24(1974), 52-66.
- [137] J.W. Meiland: The Nature of Intention, Methuen, London, 1970.
- [138] A.I. Melden: Free Action, Routledge & Kegan Paul, London, 1961.
- [139] J.S. Mill: System of Logic, University of Toronto Press, 1973.
- [140] A.R. Miller: "Wanting, Intending, and Knowing What One is Doing", PPR, 60(1980), 334-43.
- [141] T. Mischel (ed): The Self, Blackwell, Oxford, 1977.
- [142] F.C.T. Moore: The Psychological Basis of Morality, Macmillan, London & Basingstoke, 1978.
- [143] G.E. Moore: Ethics, Oxford U.P., 1912.
- [144] T. Nagel: The Possibility of Altruism, Oxford U.P., 1970.
- [145] L. Nordenfelt: Explanation of Human Action, Uppsala 1974.
- [146] R. Norman: Reasons for Action, Blackwell, Oxford, 1971.
- [147] S.E. Norris: "The Intelligibility of Practical Reasoning", APQ, 12(1975), 77-84.
- [148] P.H. Nowell-Smith: Ethics, Penguin, 1954.
- [149] P.H. Nowell-Smith: "Ifs and Cans" repr. in [76].
- [150] B. O'Shaughnessy: "Trying (as the Mental 'Pineal Gland')", JP, 70(1973), 365-86.
- [151] C. Peacocke: Holistic Explanation, Oxford U.P., 1979.
- [152] D.F. Pears: "Two Problems about Reasons for Actions" in [30].
- [153] D.F. Pears: Questions in the Philosophy of Mind, Duckworth, London, 1975.
- [154] D.F. Pears: "Ifs and Cans" repr. in [153].
- [155] D.F. Pears: "A Sketch for a Causal Theory of Wanting and Doing" in [153].
- [156] D.F. Pears: "The Appropriate Causation of Intentional Basic Actions", Critica, 7(1975), 39-69.
- [157] A. Phillips Griffiths: "On Belief" repr. in [158].
- [158] A. Phillips Griffiths (ed): Knowledge & Belief, Oxford U.P., 1967.
- [159] A. Phillips Griffiths: "Reasons and Causes" in R.F. Dearden et al. (eds): Reason, Routledge & Kegan Paul, London, 1972.
- [160] K.R. Popper: "Indeterminism in Quantum Physics and Classical Physics", The British Journal for the Philosophy of Science, 1(1950), 117-33, 173-95.
- [161] H.H. Price: Belief, Allen & Unwin, 1969.
- [162] H.A. Prichard: "Acting, Willing, Desiring" repr. in [210].
- [163] H. Putnam: Mind, Language and Reality: Phil. Papers II, Cambridge U.P., 1975.

- [164] W.V.O. Quine: Word and Object, M.I.T. Press, Cambridge Mass. 1960.
- [165] W.V.O. Quine: "Reference and Modality" in From a Logical Point of View, Harvard U.P., 2ed., 1961.
- [166] W.V.O. Quine: "Quantifiers and Propositional Attitudes" in The Ways of Paradox, Random House, New York, 1966.
- [167] J. Rachels: "Reasons for Action", CJP, 1(1971), 173-87.
- [168] C. Radford: "Hoping and Wishing", PAS, Supp. vol. 44 (1970).
- [169] J. Raz: Practical Reason & Norms, Hutchinson, London, 1975.
- [170] D.A.J. Richards: A Theory of Reasons for Action, Clarendon Press, Oxford 1971.
- [171] C. Ripley: "A Theory of Volition", APQ, 11(1974), 141-7.
- [172] G. Ryle: The Concept of Mind, Hutchinson, London, 1949.
- [173] S. Schiffer: "A Paradox of Desire", APQ, 13(1976), 195-203.
- [174] A. Schopenhauer: Die Welt als Wille und Vorstellung, Vol. I, Zürcher Ausgabe, Diogenes, 1977.
- [175] A. Schopenhauer: Ueber die Freiheit des Willens, Zürcher Ausgabe, Diogenes, 1977.
- [176] J.R. Searle: Speech Acts, Cambridge U.P., 1969.
- [177] W. Sellars: Science, Perception, and Reality, Routledge & Kegan Paul, 1963.
- [178] W. Sellars: "Some Reflections on Language Games" repr. in [177].
- [179] W. Sellars: "Empiricism and the Philosophy of Mind" repr. in [177].
- [180] W. Sellars: "Fatalism and Determinism" in [177].
- [181] W. Sellars: "Thought and Action" in [177].
- [182] W. Sellars: Science and Metaphysics, Routledge & Kegan Paul, 1968.
- [183] W. Sellars: "Language as Thought and as Communication", PPR, 29(1969), 506-27.
- [184] W. Sellars: "Actions and Events", Nous, 7(1973), 179-202.
- [185] W. Sellars: Essays in Philosophy and Its History, Reidel, Dordrecht, 1974.
- [186] W. Sellars: "Metaphysics and the Concept of a Person" repr. in [185].
- [187] W. Sellars: "Reply to Marras" repr. in [185].
- [188] W. Sellars: "Volitions Re-affirmed" in [36].
- [189] J.A. Shaffer: "Sexual Desire", JP, 75(1978), 175-89.
- [190] S. Shoemaker: "Functionalism and Qualia", PS, 27(1975), 291-315.
- [191] E. Sosa: "Propositional Attitudes De Dicto and De Re", JP, 67(1970), 883-96.
- [192] E. Sosa (ed): Causation and Conditionals, Oxford U.P., 1975.
- [193] R.C. Stalnaker: "A Theory of Conditionals" in [192].
- [194] S.P. Stich: "Do Animals Have Beliefs?", AJP, 57(1979), 15-28.
- [195] F. Stoutland: "The Logical Connection Argument", APQ Monographs, 4(1970).
- [196] F. Stoutland: "The Causal Theory of Action" in J. Manninen & R. Tuomela (eds): Essays on Explanation and Understanding, Reidel, Dordrecht, 1976.
- [197] F. Stoutland: "The Causation of Behavior" in [102].
- [198] C. Taylor: The Explanation of Behaviour, Routledge & Kegan Paul, 1964.
- [199] R. Taylor: Action & Purpose, Prentice-Hall, Englewood Cliffs, N.J., 1966.
- [200] I. Thalberg: Enigmas of Agency, Allen & Unwin, 1972.
- [201] I. Thalberg: Perception, Emotion and Action, Blackwell, Oxford, 1977.
- [202] J.J. Thomson: Acts and Other Events, Cornell U.P., Ithaca & London, 1977.
- [203] R. Trigg: Pain and Emotion, Oxford U.P., 1970.
- [204] R. Trigg: "Thought and Language", PAS, 79(1978/9).
- [205] R. Tuomela: Human Action and Its Explanation, Reidel, Dordrecht, 1977.
- [206] J.O. Urmson: "Parenthetical Verbs", Mind, 61(1952), 480-96.
- [207] Z. Vendler: Res Cogitans, Cornell U.P., Ithaca & London, 1972.
- [208] R. Ware: "Acts and Action", JP, 70(1973), 403-18.
- [209] G. Watson: "Skepticism about Weakness of Will", PR, 86(1977), 316-39.
- [210] A.R. White (ed): The Philosophy of Action, Oxford U.P., 1968.
- [211] A.R. White: Modal Thinking, Blackwell, Oxford, 1976.
- [212] C.H. Whiteley: Mind in Action, Oxford U.P., 1973.
- [213] E. Wilson: The Mental as Physical, Routledge & Kegan Paul, 1979.
- [214] J.R.S. Wilson: Emotion and Object, Cambridge U.P., 1972.

- [215] A. Woodfield: Teleology, Cambridge U.P., 1976.
 [216] G.H. von Wright: Norm & Action, Routledge & Kegan Paul, 1963.
 [217] G.H. von Wright: Explanation and Understanding, Routledge & Kegan Paul, London, 1971.
 [218] G.H. von Wright: Causality & Determinism, Columbia U.P., London & New York, 1974.
 [219] G.H. von Wright: "Freedom and Determination", Acta Philosophica Fennica, 31(1980), North-Holland.
 [220] J.J. Young: "Ifs and Hooks: A Defence of the Orthodox View", Analysis, 33(1972), 56-63.

INDEX

- Abelson, R. 116, 120, 181
 Ability 73-84 see also "Can"
 voluntary and inclusive 80-4
 Act and activity 14
 Acting knowingly 155
 Action
 basic see Basic action
 broad concept of 12-26
 counterfactual element in 20-4
 different concepts of 12-3
 intentional see Intentional
 action
 modifiers of 25
 temporal location of 17-8, 28-9
 Akrasia see Weakness of Will
 Alston, W.P. 120, 123, 180
 Analogy theory of thinking see
 Thinking
 Anscombe, G.E.M. 10, 147, 182
 Armstrong, D.M. 35-6, 42-9, 115,
 177-8, 182
 Assent see Thinking
 Assertion within the scope of
 supposition 70-3
 Audi, R. 50, 111, 120, 123-8, 169
 Aune, B. 17, 52-5, 56, 76-7, 111,
 166, 167-8, 178, 184
 Austin, J.L. 5, 70, 74, 75, 103-4,
 175
 Ayers, M.R. 79-80, 179

 Bach, K. 148-51, 184
 Background convictions 63-6
 expansion of 63ff
 Baier, A. 98, 151, 163, 180
 Baier, K. 9, 87-9, 171, 172, 175,
 179
 Basic action
 intentionality of see Inten-
 tional action)
 ontologically 26-7, 156, 183
 originally 158-63
 teleologically 140, 156, 157-63,
 183 see also Intentional action
 Beardsley, M.C. 16, 25, 134-5, 169,
 176, 183, 184
 Beebe, M. 177
 Believing 30, 34-5 see also Thinking
 Bodily actions 27, 81, 83, 158-9, 163
 and bodily processes 15-6
 Braithwaite, R.B. 46, 177-8
 Brand, M. 165, 166
 Brandt, R. 123
 Brown, D.G. 176
 Burge, T. 177

 "Can"
 and exclusion of luck 80-4
 and knowledge how to act 82-4,
 152-5
 and repeatability of performance
 81-2
 and skill 81-4, 155
 general function of 84-5
 hypothetical analysis of 73-80
 objections to hypothetical analysis
 of 76-80
 of ability and opportunity 73-84
 summary of senses of 84
 Cancellability (of implications) 112
 Castañeda, H-N. 166-7, 184
 Causal chains
 non-standard external 152-5, 160-3
 non-standard internal 135ff
 Causal theory of mind 42-51
 Causation 22-5, 182
 and inus-conditions 23-4
 and triggering conditions 24
 immanent 18
 immediate knowledge of 120-2
 independence of action 18-22
 overdetermination 23-4
 relata of 22-3
 transitivity of 24-5

- Circumstantial action 27-8, 66, 160-2
- Chisholm, R.M. 18, 63, 67, 76-7, 142, 157, 177, 178, 184
- Churchland, P.M. 117, 169
- Complex action
 ontologically 29
 teleologically 151-2
- Conditionals 63-73
 Austinian 70-3
 contraposition of 68
 counterfactual 69-70
 negation of 66
 open 63-9, 70-3
 practical 70-3
 presuppositions of 66-7, 77-8
 reductio-employment of 68-9
 transitivity of 69
- Confidence 32
- Consequential action 28-9, 160-2, 184
- Consideration-making belief 87-9
- Danto, A.C. 26, 111, 142, 176, 183
- Davidson, D. 17, 22, 25, 26, 65, 76-7, 90, 100, 110, 122, 135-6, 152, 157, 172, 175, 176, 180, 181, 182, 185
- Davidson's mountaineering-example 135ff, 144-5, 166-7
- Davis, L.H. 17, 166, 183
- Decision 10, 32, 166, 169-70
- Dennett, D. 178
- Descartes, R. 32, 33
- Desire, sexual 95-6 see also Wanting
- Disposition, realist view of 35-6, 44, 115-6
- Edgley, R. 130-2, 177, 183
- Emotion 32, 48-9, 105-6, 123-8, 173-4
- Ends, justification of 171-2
- Epistemic possibility see Possibility
- Essential action 28-9
- Events 22
- Existential generalization 37-40
- Explicit logical relations 35, 41-2
- Facts 22
- Feed-back 141-2, 145-7, 148
- Feelings see Emotion and Wanting
- Fodor, J.A. 29, 61, 176, 177, 178
- Foley, R. 136-7
- Frankfurt, H.G. 98, 138, 139, 140, 145-6, 183
- Functionalism see Causal theory of mind
- Ginsberg, M. 50
- Goldman, A.I. 27-8, 90, 111, 117, 137-8, 153, 157, 166, 175, 179, 181-2, 183
- Goodman, N. 70
- Grice, H.P. 178
- Habitual action 160
- Hampshire, S. 104, 111
- Harman, G. 180, 185
- Hornsby, J. 17, 181, 183
- Hume, D. 32, 101, 120, 180-1
- Hunter, J.F.M. 177
- Hypothetical analysis of "can" see "Can"
- Ideomotor theory 184
- "If" 63-73
- Images (isomorphic symbols) 56, 59-62, 148-50
- Immanent causation see Causation
- Inner speech see Thinking
- Intensional contexts see Thinking
- Intention 166, 169-70, 177, 180 and prediction 170
- Intentional action 134-63
 basic 134-51, 157-63
 causal accounts of 135-63
 complex 151-2
 intensional context of 143-4
 mental 151
 non-basic 152-7
 non-causal accounts of 20-1, 134-5, 138-9, 146
- Intentional system 178
- Intransitive verbs see Transitive verbs
- Intrinsic result 14-5
- Inus-condition see Causation
- Isomorphic symbols see Images
- James, W. 146-7, 150-1, 184
- Kant, I. 111, 155, 181
- Kaplan, D. 39
- Kenny, A.J.P. 96, 102, 115, 166, 175
- Kim, J. 123
- Knowledge how to act see "Can"
- Kripke, S.A. 40, 42, 177
- Language and reality 5
- Language-learning 5, 57-9
- Lehrer, K. 76, 77-8
- Levin, M. 177
- Linguistic phenomenology 5

- Locke, D. 114-5, 116, 175, 179
- McCann, H. 34, 166, 183
- McCullagh, C.B. 183
- Mackie, J.L. 23, 63, 67, 68, 70-2, 176, 182
- Manipulative action 27, 160-2
- Manser, A. 175
- Marras, A. 178
- Martin, R. 183
- Means and conditions 83, 155-7
- Meiland, J.W. 180
- Melden, A.I. 16
- Mill, J.S. 176
- Moore, G.E. 75
- Motivating circumstance 10
- Nagel, T. 133-4
- Non-basic action see Intentional action and Basic action
- Nordenfelt, L. 182
- Norman, R. 89-90
- Normativity 41-2, 52-5, 89-90, 177
- Norris, S.E. 114
- Nowell-Smith, P.H. 181
- "Only if" 73
- Ontic possibility see Possibility
- Ordinary language, reliability of 4-5
- O'Shaughnessy, B. 166, 183
- Ought-to-be and ought-to-do 54-5, 58
- Pain 106, 107-8, 126-7, 173
- Parentetical use of psychological verbs 176-7
- Pears, D.F. 117-20, 183, 184
- Phenomenal reasons see Reasons
- Phillips Griffiths, A. 90, 179
- Philosophical method 4-7
- Possibility
 absolute and relative 63-4
 conditional ontic 74-6
 epistemic 64, 77-9
 ontic 64
- Practical knowledge 147-8
- Practical modality see Volition
- Practical reasoning 92-3, 113-4 129-34, 155-7
- Predictions, limitations of 78-9
- Price, H.H. 30, 32, 48-9
- Prichard, H.A. 17, 98, 180
- Propositions 9-10, 41-2
- Putnam, H. 177
- Quine, V.W.O. 37, 39
- Rachels, J. 175
- Radford, C. 181
- Rationality
 principles of 132
- Raz, J. 100
- Reason-governed action 1, 129, 157
- Reasons for action
 and causes 110-1
 and for belief 3-4, 8-9, 91-3
 and for wanting 91-2, 110-2
 conditional form of 10-1, 92, 155-7
 definition of 93
 dependence on wanting 86-91
 in evaluative and explanatory contexts 8-9, 90-1, 131-4, 171-2
 phenomenal and real 8-9, 93, 171-2
 second-order 100-1
- Recipes 82-4, 152-5
 performance and acquisition 158-9
- Reflexes 95-6, 106, 107-8, 126-7, 149
- Representations
 effective and receptive 148-51
- Richards, D.A.J. 132-3
- Rigid designation 42
- Ryle, G. 34, 35, 47, 66, 84, 178
- Schopenhauer, A. 182
- Self-deception 178
- Sellars, W. 34, 36-7, 52-5, 58-9, 166, 169, 178, 184
- Shaffer, J. 95-6
- Simple action 31, 152
- Sosa, E. 177
- Stalnaker, R.C. 66, 68, 69
- States of affairs see Propositions
- Stoutland, F. 20, 116, 138-9, 183
- Taylor, C. 116
- Taylor, R. 18, 142, 183, 184
- Teleological action 157
 see also Intentional action
- Tendency 109, 116
- Thalberg, I. 15, 28, 136
- Theoretical concepts 50-1, 123-8
- Theoretical reasoning 41-2, 92-3, 184
- Thinking, propositional
 analogy theory of 52-5
 and assent 32
 and behaviour 42-51
 and believing 30-5
 and emotions 48-9
 and entertaining a proposition 30-2, 45-9
 and images see Images
 and inner speech 33
 animal 56-7
 as requiring a code 33

- causal theory of 42-51
- de dicto and de re 37-40
- independence of language 56-62
- intensional context of 36-41
- knowledge of 50-1
- not an act 33-4
- reflective and unreflective 31
- relation to wanting 47-8, 50
- theoretical construct view of 49-51
- Thomson, J.J. 176
- Transitive verbs 13, 15, 24-5
- Trying 139-40, 142-3
- Tuomela, R. 26, 123, 166, 178
- Vendler, Z. 33, 177
- Volitions 149-50, 166-8
 - causal impotence of 166-8
- Von Wright, G.H. 14, 18-22, 23, 111, 121, 138, 139, 156, 176
- Wanting
 - and ability 102ff
 - and being forced 111-2
 - and emotion 123-8
 - and feelings 109
- and means 111-2, 113-4, 129-34, 135-6, 155-7
- and wishing 93-4, 104-5, 185
- as a causally explanatory concept 117-20
- as an internal state 114-20
- based on reasons 110-2
- dispositional analysis of 101-28
- dominant 112
- extensional and intensional 95-6, 106, 107-9
- knowledge of 120-3
- object of 93-100
- relation to reasons 86-91
- relation to thinking 48-9, 50
- second-order 97-100
- theoretical construct approach to 123-8
- thought-dependence of
 - manifestation of 107-9, 140ff
- Weakness of will 131-3, 172-3
- Whiteley, C.H. 107, 166
- Wilson, J.R.S. 184
- Young, J.J. 179

LUND MONOGRAPHS IN PRACTICAL PHILOSOPHY

Edited by Göran Hermerén

The studies in this series discuss assumptions, problems, methods and theories in practical philosophy. The term »practical philosophy» is then used to refer to several philosophical disciplines, including ethics, meta-ethics, theory of value, aesthetics, philosophy of action, political philosophy, philosophy of education, philosophy of religion, philosophy of law, etc. Historic as well as systematic studies in philosophy will be published in this series.

Subscriptions to the series and orders for single volumes should be addressed to any international bookseller. Universities, libraries, learned societies and publishers of learned periodicals may obtain the volumes of the series in exchange for their own publications. Inquiries should be addressed to Universitetsbiblioteket, Box 1010, S-221 03 Lund, Sweden, or to Filosofiska Institutionen, Kungshuset, Lundagård, S-223 50 Lund, Sweden.

1. Ingrid Petersson. *Utilitarianism, Responsibility and Punishment, with Special Reference to R.B. Brandt's Defence of Utilitarianism*, Lund, 1976, 215 pp. Price Sw. Cr. 25:—
2. Tore Nilstun, *Moral Reasoning — A Study in the Moral Philosophy of S.E. Toulmin*, Lund 1979, 150 pp. Price Sw. Cr. 20:—
3. Tore Nilstun, *Aristotle on Freedom and Punishment*, Lund 1980, 87 pp.
4. Ingmar Persson, *Reasons and Reason-governed Actions*, Lund, 1981, 200 pp.